VŠB – Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra informatiky

Detekce obličejových bodů v prostředí automobilu
Facial Landmarks Detection in Vehicles

2020

Petr Krumpolc

VŠB - Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra informatiky

Zadání bakalářské práce

Student: **Petr Krumpolc**

Studijní program: B2647 Informační a komunikační technologie

Studijní obor: 2612R025 Informatika a výpočetní technika

Téma: Detekce obličejových bodů v prostředí automobilu
Facial Landmarks Detection in Vehicles

Jazyk vypracování: čeština

Zásady pro vypracování:

Detekce obličeje a jeho částí je důležitým krokem v mnoha oblastech analýzy obrazu a slouží například k rozpoznání únavy, emocí, či zdravotních problémů. V posledních letech se v této oblasti využívají tzv. landmarky, což jsou body reprezentující jednotlivé části obličeje a které jsou mapovány na obličej v obraze. Cílem práce je vyzkoušet různé existující algoritmy detekce obličejových bodů v obrazech a porovnat jejich úspěšnost v reálných podmínkách v automobilu, kdy se hlava řidiče otáčí a navíc se mění světelné podmínky.

Ve své práci proveďte:
1. Představte existující vybrané metody detekce obličejových bodů.
2. Otestujte metody na videosekvencích pořízených v reálných podmínkách v automobilu.
3. Zhodnoťte úspěšnost porovnaných metod.

Seznam doporučené odborné literatury:

[1] Wu, W.,Qian, Ch., Yang, S., Wang, Q., Cai, Y., Zhou, Q.: Look at Boundary: A Boundary-Aware Face Alignment Algorithm, CVPR 2018
[2] Cao, Z., Simon, T., Wei, S-E., Sheikh, Y.: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, CVPR 2017
[3] Kazemi, V., Sullivan, J.: One Millisecond Face Alignment with an Ensemble of Regression Trees, CVPR 2014

Formální náležitosti a rozsah bakalářské práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí bakalářské práce: **Ing. Michael Holuša, Ph.D.**

Datum zadání: 01.09.2019
Datum odevzdání: 30.04.2020

VYSOKÁ ŠKOLA BÁŇSKÁ-TECHNICKÁ UNIVERZITA * Fakulta elektrotechniky a informatiky * ID16 * OSTRAVA *

doc. Ing. Jan Platoš, Ph.D.
vedoucí katedry

prof. Ing. Pavel Brandštetter, CSc.
děkan fakulty

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Ostravě 30.04.2020 .....

Děkuji vedoucímu práce Ing. Michaelu Holušovi, PhD za rady, vedení práce, a v neposlední řadě i za poskytnuté laboratorní videomateriály k detekci obličejových bodů.

Můj dík patří i mému tátovi za pomoc a ochotu při pořizování videosekvencí v prostředí automobilu.

Abstrakt

Bakalářská práce se zabývá detekcí, respektive predikcí, obličejových bodů ve videosekvencích pomocí technik strojového učení. Práce je věnována existujícím metodám detekce obličejových bodů, jejichž implementace jsou volně dostupné na internetu. Jsou prezentovány zvolené metody a jejich úspěšnost na vlastních videosekvencích, pořízených v reálných podmínkách automobilu. Úspěšností se rozumí přesnost souřadnic predikovaných obličejových bodů vůči ručně zadaným souřadnicím korespondujících bodů.

Klíčová slova: Počítačové vidění; Strojové učení; Detekce obličejových bodů; Neuronové sítě; Evaluace metod

Abstract

The proposed bachelor thesis is focused on detections, respectively predictions of facial landmarks in video sequences by using the technics of machine learning. The thesis is dedicated to existing methods of facial landmark detections whose implementations are freely available on the internet. The success rate of proposed methods which have been tested on the custom video sequences taken in the real-world conditions of traffic are presented. With the term of success rate is meant the accuracy of predicted facial landmarks against to manually annotated corresponding landmarks.

Key Words: Computer Vision; Machine learning; Facial landmark detections; Neural networks; Model evaluation

Obsah

Seznam použitých symbolů a zkratek

Seznam ilustrací

Seznam tabulek

1. Úvod

Zpracování obrazu lidské tváře prochází v současné době poměrně dynamickým vývojem, stejně tak, jako celá věda počítačového vidění. Zejména aplikace hlubokého učení na problémy počítačového vidění představuje významný milník. Jednou z oblastí aplikace hlubokého učení je v současnosti predikce obličejových bodů. Koncept hlubokého učení není nový, avšak až díky současnému výpočetnímu výkonu je možné hluboké učení stále více užívat v praxi. Nicméně jak tato práce později ukáže, existují i jiné nástroje umělé inteligence, schopné predikce.

Všechny zvolené metody predikce obličejových bodů charakterizované v této práci, jsou postaveny na technikách strojového učení, respektive hlubokého učení a rozhodovacích stromů. Proto se ani tato práce, v teoretické rovině, těmto tématům nevyhne.

K čemu je vlastně potřeba predikovat body na tváři člověka? Nabízí se několik možných aplikací. Jsou to například bezpečností systémy. Detekce obličejových bodů je jednou z fází rozpoznání identity člověka. Lze také rozpoznávat náladu člověka na základě jeho výrazu v obličeji. Zajímavá aplikace může být rovněž v chytrých telefonech, kdy za pomocí předního fotoaparátu lze rozpoznat, zda uživatel sleduje displej telefonu a zabránit tak zhasnutí displeje. V automobilech by takový systém mohl rozpoznat únavu, hněv nebo obecně emoce řidiče. V těchto případech má své místo právě detekce obličejových bodů.

Ačkoliv se tato práce nezaměřuje na konkrétní aplikaci, jako spíše na testování, závěry této práce mohou nalézt využití v praktické implementaci.

Predikci obličejových bodů může doprovázet celá řada problémů ústící v nesprávný výstup. Jestliže je predikce jedním z kroků komplexního řešení, zvyšuje se riziko, že následující fáze obdrží nesprávná data, tj. chybnou predikci. Selže-li detekce obličeje, fáze obvykle předcházející predikci obličejových bodů, selže i samotná predikce. Chyba je tak dále distribuována napříč řešením. Příčiny chybné predikce obličejových bodů mohou být např. světelné podmínky, částečné zakrytí tváře, mimika tváře, natočení obličeje atp. Selhání může samozřejmě plynout i z technického řešení, např. nedostatečně natrénovaný model neuronové sítě nebo nezvládnutá předpříprava zpracování obrazu. Nevhodně zvolené technologické řešení k uspokojivým výsledkům také nepovede.

Cílem práce je otestovat funkčnost zvolených metod na snímcích z prostředí automobilu, proto je součástí této práce rovněž datová množina pořízená v interiéru vozidla, v reálném silničním provozu a za různých světelných podmínek, kdy se hlava řidiče otáčí. Na úvod lze zmínit fakt, že žádná z testovaných metod nebyla trénována na obličejích v prostředí automobilu. Je tedy otázkou, jak obstojí zvolené metody na testovací datové množině.

Zvolené metody jsou testovány na datové množině z reálných podmínek automobilu i na materiálech z laboratorních (kontrolovaných) podmínek. Vybrané metody jsou čtenáři nejprve představeny. Závěr práce se věnuje testování zvolených metod.

2. Počítačové vidění

Jak uvádí autor v knize [1] *„Počítačové vidění spadá do oblasti počítačové vědy, matematiky a elektrotechniky. Zabývá se analýzou, zpracováním, získáním a porozuměním obrazovým datům ze skutečného světa s cílem napodobit lidské vidění. Na rozdíl od lidského vidění, je možné počítačové vidění využít k analýze a zpracování hloubkových a infračervených snímků.“* Jde tedy o poměrně široce zaměřenou vědní disciplínu, byť snahou je pouze přiblížení se ke schopnostem lidského zraku. Použitím speciální techniky je možné člověka v některých oblastech i překonat, např. analýzou snímků pořízených mimo viditelné spektrum člověka.

Podle autora knihy [1] patří mezi úlohy této disciplíny např.:

- klasifikace, detekce a rozpoznání objektu,
- analýza a detekce pohybu,
- rekonstrukce scény a obrazu.

Spolu s rozvojem strojového učení, a zejména hlubokého učení, došlo postupem času k začlenění této techniky i do problematiky počítačového vidění. Možnosti zpracování lidské tváře, respektive detekce obličejových bodů, jsou stále zlepšovány právě díky použití technik strojového učení.

Následující podkapitoly jsou věnovány obrazovým datům, respektive jejich paměťové reprezentaci a anomáliím spojených s jejich pořizováním. V podkapitole 2.2 jsou velmi stručně představeny knihovny usnadňující práci s obrazem.

2.1. Obrazová data

Počítačové vidění pracuje s videosekvencemi, respektive s jednotlivými obrazovými snímky. Ty mohou být pořízený různými typy snímačů. V této práci byl k získání videí z reálných podmínek automobilu využit digitální fotoaparát, který je schopen zaznamenávat i video. Byla využita i kamera, a to k pořízení snímků z laboratoře. Získaným snímků se podrobněji věnuje kapitola 4.2.

2.1.1. Paměťová reprezentace

V paměti počítače jsou obrazová data nejčastěji reprezentována formou dvoudimenzionální matice pixelů v příslušném barevném modelu. V rámci této práce jde o RGB, respektive BGR a obrázek v odstínech šedi.

2.1.2. Pořizování obrazových dat

Pokud lze ovlivnit podmínky, za kterých jsou data získávána, lze zvýšit nejen úspěšnost řešení, ale také jeho efektivitu nebo jednoduchost.

Při detekci obličeje řidiče je výhodou statická kamera. Neměl by tedy být problém kameru správně zaostřit na požadovanou vzdálenost, již v době pořizování videosekvencí.

Komplikovanější situace nastává při zajištění ideálních světelných podmínek, respektive minimalizaci zejména ostrých stínů. V reálných podmínkách automobilu mohou vznikat velice snadno, a to zejména po setmění. Zdrojem pak mohou být pouliční lampy, světla protijedoucích vozidel nebo pouhé světlo v interiéru vozidla.

E.R. Davis [2] vnímá stíny jako zdroj problémů analýzy obrazu. Jednak je problematické stíny detekovat, ale také přidávají další hrany, které musí být zpracovány při detekování objektů. Eliminace stínu může být zajištěna *„[…] snížením jejich kontrastu použitím několika zdrojů světel. Potom se oblast stínu jednoho zdroje světla stane osvětlenou oblastí díky jinému zdroji světla, čímž se dramaticky sníží kontrast stínu. Ve skutečnosti, pokud existuje* n *světel, mnoho zastíněných míst bude osvětleno* $n-1$ *světly a jejich kontrast bude tak malý, že budou moci být eliminovány prostým prahováním.“* [2]

V nekontrolovatelných podmínkách, jako je např. automobil, tento typ řešení naneštěstí nelze použít. K uvedenému řešení by mohlo ale dojít náhodou, kdyby na obraz přirozeně působilo více světel dohromady, např. více pouličních lamp. Nicméně takové řešení samozřejmě není stabilní. Hypoteticky, použití nějakých přídavných světel, které by správně osvětlovaly hlavu řidiče, také nelze využít. Zvlášť za jízdy ve tmě by v interiéru vozidla bylo více světla než v exteriéru vozu, čímž by došlo k podpoření odrazů v oknech vozidla. Jinými slovy, bylo by velice obtížné pro řidiče sledovat okolí vně vozu.

K možnému řešení tohoto problému by mohlo vést použití kamery s nočním viděním nebo kombinací klasické kamery a termokamery. Případně použitím více kamer, snímající řidiče z různých úhlů v jeden okamžik. Čímž by samozřejmě došlo k rapidnímu zvýšení nároků na výpočetní výkon, ale zároveň by se zvýšila šance, že se alespoň obraz z některé z kamer podaří úspěšně analyzovat.

2.2. Knihovny

V praktické části této práce jsou pro usnadnění programování využity knihovny pro základní práci s obrazem.

První využitou knihovnou je **OpenCV**. Jedná se o multiplatformní řešení jak pro počítačové vidění, tak pro strojové učení. Sestává se z komplexní nabídky algoritmů pro úlohy jako detekce a rozpoznání tváře, sledování trajektorie pohybu objektů, klasifikaci objektů apod. [3]

Druhou využitou knihovnou je **Dlib**. Znovu jde o multiplatformní řešení s otevřeným zdrojovým kódem, vyvinuté v programovacím jazyce C++. Knihovna je kolekcí softwarově nezávislých komponent. Konkrétně jde o komponenty např. pro data mining, zpracování obrazu, strojové učení apod. [4]

3. Strojové učení

Strojové učení spadá do oblasti umělé inteligence a jeho aplikace bychom mohli hledat ve zpracování lidské řeči, předpovídaní vývoje burzy nebo právě při detekci obličejových bodů.

Algoritmy strojového učení fungují v podstatě na bázi extrahování informací z dat, a jsou reprezentovány příslušným modelem. Model pak slouží ke zpracování dat, které nebyly použity při jeho tvorbě. [5]

V této práci budou představeny modely typu:

- neuronových sítí,
- rozhodovacích stromů.

Rovněž všechny modely jsou trénovány pomocí metodiky Učení s učitelem (Supervised learning), což znamená, že trénovací množina je anotovaná, respektive jsou známy vstupy a očekávané výstupy. Model se tedy snaží najít vhodnou aproximační funkci, mapující vstupy na očekávané výstupy.

3.1. Neuronové sítě

Neuronová síť je síť propojených neuronů, která však zatím není zdaleka tak rozsáhlá jako je tomu u lidí.

Učení neuronových sítí probíhá primárně na základě změn příslušných vah neuronu. Váha určuje důležitost určitého spojení neuronu, respektive relevantnost informace daného propojení. Pokud je váha 0, je tento vstup ignorován. Analogicky se v případě nenulové váhy, vstup dostane ke zpracování výstupní aktivační funkcí. Mezi algoritmy učení neuronových sítí lze zařadit Back Propagation. Tento algoritmus se snaží, na základě výstupu sítě, minimalizovat chybu upravováním příslušných vah. [5]

Chování neuronových sítí udává jejich architektura, respektive počet neuronů, počet vrstev neuronů a typy spojení mezi vrstvami.

3.1.1. Aktivační funkce

Aktivační funkce popisuje transformaci výstupní hodnoty neuronu. Nejtriviálnější aktivační funkcí je pravděpodobně lineární (jednotková) aktivační funkce, která propouští nepozměněný signál ven z neuronu. Nejčastěji se tato funkce používá ve vstupní vrstvě neuronových sítí.

Další užitečnou aktivační funkcí je Sigmoid, která umožňuje normalizovat dat do rozmezí hodnot 0,0–1,0. Normalizaci dat do rozmezí od -1,0 do +1,0 zajišťuje hyperbolická funkce TanH.

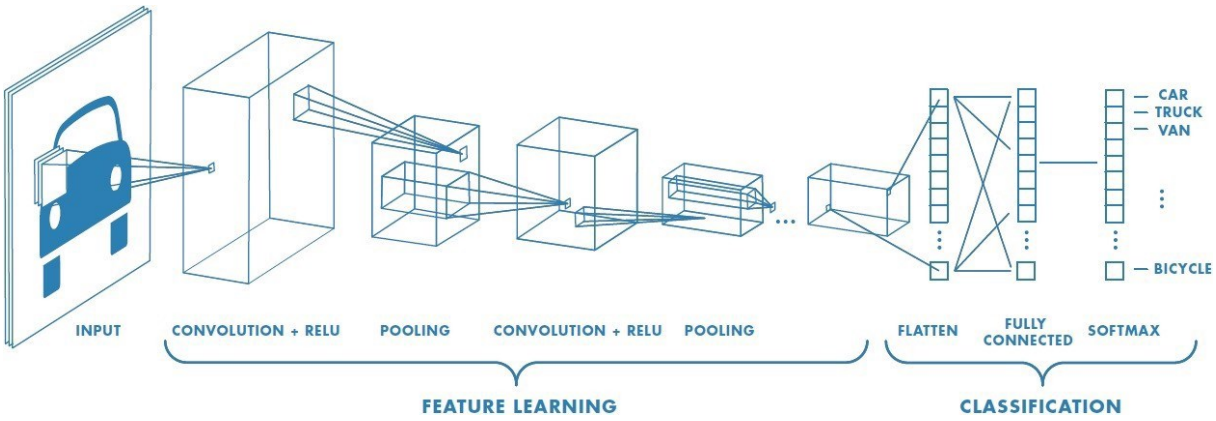
V případě potřeby ořezu záporných hodnot, poslouží aktivační funkce ReLU (Rectified linear units). Pokud je výstup agregační funkce záporný, je výstupem neuronu hodnota 0. V případě kladného výstupu agregační funkce, se aktivační funkce chová lineárně a nedochází k modifikaci hodnoty. [5]

3.2. Konvoluční neuronové sítě

Jde o typ neuronové sítě vhodný, mimo jiné, ke zpracování obrazových dat. Důvodem je dobrá škálovatelnost sítě, kdy narozdíl od plně propojené neuronové sítě jsou neurony jedné vrstvy propojeny jen se skupinou neuronů předchozí vrstvy. Tímto dojde k redukci spojů mezi neurony. Dalším důvodem je efektivita zpracování dat mající strukturu (skupina sousedících pixelů tvořící vzor). [5]

Kromě vstupní a výstupní vrstvy neuronů je konvoluční neuronová síť také tvořena skrytými vrstvami. Tyto skryté vrstvy mají za úkol extrahovat příznaky z dat. Mohou být tvořeny opakováním vrstev:

- konvoluce (convolution),
- sdružení (Pooling).



Obrázek 1 – Architektura klasifikační konvoluční neuronové sítě. [25]

Příklad konvoluční neuronové sítě znázorňuje Obrázek 1. Síť je rozdělena na dvě podsítě, respektive na podsíť učení/extrahování příznaků (Feature learning) a na podsíť klasifikace (Classification). Klasifikační podsíť je plně propojená neuronová síť jejíž výstupní neurony zařazují vstupní snímek do správné kategorie (auto, nákladní auto, dodávka atd.). Podsíť extrahující příznaky je tvořena vrstvami blíže popsanými v následujících podkapitolách.

3.2.1. Konvoluční vrstva

Konvoluční vrstva (Obrázek 1 Convolution + ReLU) je konvoluční filtr, který je aplikován na vstupy neuronu. Respektive samotné váhy vstupů jsou koeficienty konvolučního filtru. Zpracování si lze představit jako vstupní matici dat o dané velikosti, na niž je aplikován filtr (představován maticí), který je zpravidla menší než vstupní matice. Tento filtr je posouván po vstupní matici a produkuje novou matici příznaků. [5]

3.2.2. Sdružující vrstva

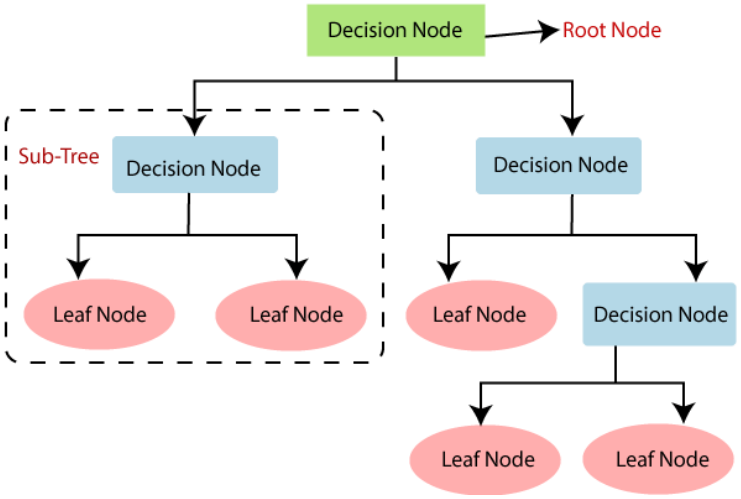
Tento typ vrstvy (Obrázek 1 Pooling) se typicky vkládá mezi konvoluční vrstvy za účelem redukce velikosti (šířky a výšky) vstupních dat. [5]

3.3. Rozhodovací stromy

Rozhodovací stromy jsou alternativou k neuronovým sítím. Mohou být použity nejen ke klasifikaci, ale i k regresi neboli predikci.

Výhodou rozhodovacích stromů je snadná interpretace a testovatelnost modelu. Protože je možné odpozorovat, jak model funguje, a to včetně kroků algoritmu během jeho tvorby, je možné využít tzv. white-box testování[1]. Plusem jsou také menší nároky na výpočetní výkon. Nevýhodou rozhodovacích stromů někdy bývá až zbytečně komplexní model, který ovšem vznikl na základě trénovacích dat. [6]

Vizualizovat rozhodovací strom, lze jak jinak než právě stromovou strukturou, viz Obrázek 2. Strom je klasicky tvořen z vrcholů a hran. Každý vrchol (mimo listy) představuje větvení [6]. Jde tedy o pravidla, která je možné interpretovat jako podmínky if – else, známé z mnoha programovacích jazyků.



Obrázek 2 – Vizualizace obecného rozhodovacího stromu.
Zeleně – kořenový uzel. Modře – rozhodovací uzly. Červeně – listy. [21]

[1] Jedná se o protipól k black-box testování, u kterého není známa struktura kódu nebo způsob nakládání s daty. U white-box testování vývojář zná kód, způsob zpracování dat a může přesněji otestovat slabiny takového řešení.

4. Datové množiny

Než bude přistoupeno k představení zvolených metod detekce obličejových bodů, bude tato kapitola věnována ke stručnému představení datových množin, jež jsou relevantní k této práci.

Datovou množinou, ve smyslu detekce obličejových bodů, je myšlena sada obrazových podkladů tváří lidí, zachycených v různých podmínkách. Tyto datové množiny jsou jednak využívány k **trénování** modelů strojového učení k detekci obličejových bodů, jsou ale také využívány k **testování** již natrénovaných modelů. Čím jsou parametry datových množin lepší, tím mohou být lepší i predikce obličejových bodů.

V podkapitole 4.1 jsou stručně charakterizovány datové množiny, jež byly využity autory metod k natrénování modelů detekce obličejových bodů (metody budou představeny v kapitole 5).

Žádná z metod nebyla natrénována na obličejích v prostředí automobilu, a právě proto je testování zaměřeno na ověření funkčnosti metod pro snímky z prostředí automobilu. Z tohoto důvodu byla pořízena testovací datová množina z prostředí interiéru vozidla, zachycující řidiče vozidla (viz kapitola 4.2.2). Pro doplnění snímků z reálných podmínek zejména o mimiku a gesta, je v kapitole 4.2.1 představena datová množina pořízená v laboratorních podmínkách.

4.1. Trénovací datové množiny

Jak již bylo zmíněno v úvodu, jde o sady snímků použitých pro trénování modelů metod detekce obličejových bodů. Jde o:

a) Trénovací množina **300 Faces in-the-Wild (300W)** je tvořena spojením existujících datových množin (LFPW, AFW – Obrázek 3d, Helen, XM2VTS a iBug – Obrázek 3c). Tváře jsou anotovány polo-automaticky 68 obličejovými body [7]. Celkem je k dispozici $\sim 4\,000$ snímků, obsahující tváře spíše z čelního pohledu. [8]
b) Trénovací datová množina **Helen** je tvořena celkem $2\,000$ obrázky s vysokým rozlišením. Ukázkou jednoho snímku z množiny je Obrázek 3a. Tváře na snímcích jsou větší než $500\ px$ na šířku. Snímky zachycují různé pózy, výrazy nebo světelné podmínky. Neobsahují tváře zachycené z profilu. Obličeje jsou anotovány 194 obličejovými body. [9]
c) **Wider Facial Landmarks in-the-wild (WFLW)** je datová množina čítající $7\,500$ tváří pro trénování modelů. Je pořízena v nekontrolovaných podmínkách, viz Obrázek 3b. Důraz je kladen zejména na velkou variabilitu póz, výrazů a zakrytí tváře. Tváře jsou manuálně anotovány 98 obličejovými body. [10]
d) **300W-LP** je uměle vygenerovaná datové množina, získána vykreslením tváří z množiny 300-W v rozsahu natočení hlavy od $-90°$ do $+90°$. Obsahuje $61\,225$ obrázků s 2D i 3D anotacemi obličejových bodů. [8] Ukázkou vygenerovaného snímku je Obrázek 3e.

(a) (b) (c) (d) (e)

Obrázek 3 – Ukázky snímků z trénovacích množin.
(a) – Helen [22]; (b) – WFLW [24]; (c) – iBug [23]; (d) – AFW [23];
(e) – LFPW – vygenerován pro 300W-LP na základě snímku z LFPW [23]

4.2. Testovací datová množina

Specificky pro účely bakalářské práce bylo pořízeno množství videosekvencí zachycující řidiče v automobilu. Podařilo se dosáhnout poměrně rozsáhlé variability videosekvencí v nichž se střídají světelné podmínky, zachycené osoby, gesta, mimika, pozadí atd. Byla pořízena laboratorní videa, která jsou bohatá zejména na mimiku obličeje, i videa z reálného provozu. Většina videí snímá pouze řidiče, ovšem za různých podmínek. Rozlišení videí, respektive obrazových sekvencí je 1280×720.

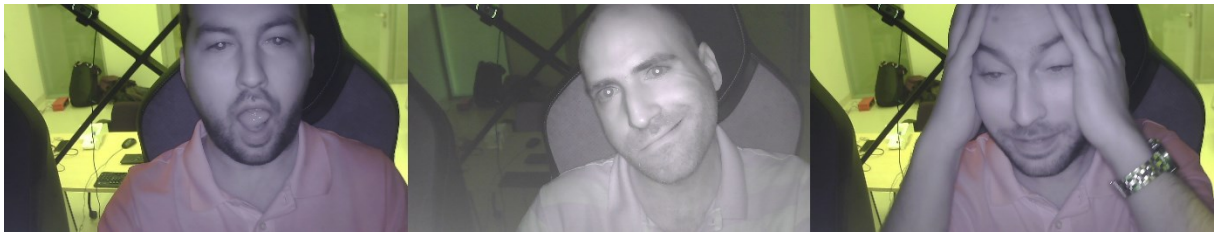
4.2.1. Laboratorní podmínky

Materiály patřící do této skupiny vznikly ve velké míře uměle v laboratorních podmínkách. Jedná se o skupinu videomateriálů převzatých od vedoucího práce. Velkou výhodou těchto materiálů je důraz na prvky, jež by mohly ústit v neúspěch metod. Testují tak krajní případy.

Kamera je nejčastěji umístěna přímo před řidičem nabízející čelní pohled. Tento případ je asi nejideálnější, jelikož nabízí dobrý záběr na řidiče, a to i v případě otáčení hlavy. V jednom případě je kamera umístěna na A – Sloupku, zatímco řidič se nachází na pravé straně vozu (vozidlo se řízením vpravo). Nevýhodou tohoto umístění je ve výsledku jen malá oblast, ve které se nachází obličej řidiče. Což komplikuje detekci obličeje, která je často nutnou součástí samotné predikce obličejových bodů.

V materiálech se střídají celkem 3 osoby řidičů. Z toho jsou 2 muži a 1 žena. Dvě videa jsou pořízena v laboratoři a jedno v uměle vytvořeném kokpitu automobilu. Pozadí je většinou statické, případné změny jsou zapříčiněné pohybem jiných osob v laboratoři. Obrázek 4 obsahuje ukázky tří snímků z laboratorních podmínek.

Doménou těchto videomateriálů je simulování různých typů gest a mimiky. Nabízí se kašel, silný kašel, mluva, zavíraní očí, klimbání hlavy, rozhlížení do stran, zívání, úsměv, neutrální výraz, ruka před ústy, škrábání se na krku, mnutí očí, upravování se, nebo úplné zmizení ze záběru (např. úpadek do mikro spánku).



Obrázek 4 – Ukázka datové množiny z laboratoře.

4.2.2. Reálné podmínky

Do této kategorie spadají vlastní videomateriály pořízené v automobilu za reálného provozu. Očekávaným efektem měl být fakt, že řidič se plně soustředí na řízení a všechna jeho gesta jsou automatická a odpovídají aktuální dopravní situaci. Možnou nevýhodou je, že materiály nedisponují tak pestrou škálou mimiky a gest, které by případně mohly vést k selhání testované metody.

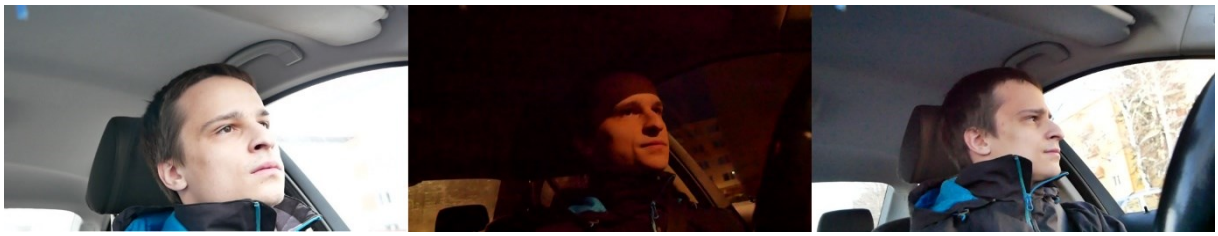
Kamera snímá řidiče z pozice středové konzole ve výšce průduchů klimatizace. Výhodou takového umístění je dobrý výhled na řidiče. Obličej řidiče zabírá většinu plochy záběru, což přispívá k lepším výsledkům detekce obličeje. Avšak nevýhodou je, že součástí záběru je i řidičovo boční okno. Kamera tak může být zaslepena sluncem nebo jiným zdrojem světla. Pokud je hlava řidiče natočena směrem k oknu, může dojít k zániku hrany obličeje právě vlivem světelných podmínek. Zvyšuje se rovněž pravděpodobnost detekce, respektive predikce obličejových bodů osob mimo vozidlo.

Videomateriály zachycují pouze jednoho řidiče, v jednom typu vozidla, povětšinou s neutrálním výrazem na tváři. Viz Obrázek 5. Časté je různé naklánění a otáčení hlavy dle dopravní situace. S tím se pojí i příslušná gesta a emoce. Část tváře je často kryta límcem zimní bundy, někdy na okamžik i rukou.

Vznikly celkem tři videomateriály. Každý za jiných světelných podmínek.

Večer za tmy, kdy do interiéru vozu proniká okolní světlo z pouličních lamp. Video je díky tomu bohaté na stíny, kamera se tak musí často přizpůsobovat na rychle se měnící světelné podmínky. Tyto záběry jsou díky zvýšené ISO hodnotě snímače poměrně šumové.

Zbylé dva materiály vznikly ve dne. V jednom případě za jasna, ve druhém bylo zataženo. Světlo za zatažena je všesměrové a vytváří spíše měkké stíny, díky čemuž ovlivňuje nejméně obraz.



Obrázek 5 – Ukázka datové množiny z prostředí automobilu.
Levý snímek – zataženo. Prostřední snímek – noc. Pravý snímek – jasno.

5. Detekce obličejových bodů

Na detekci obličejových bodů lze také pohlížet jako na problém získávání geometrie neboli struktury lidské tváře z obrazu. Tato geometrie je získávána jako vektor 2D souřadnic, představující jednotlivé body, jež jsou mapovány na lidský obličej. Každý takový bod je asociován s určitou pozicí na tváři, např. pravý koutek úst, špička nosu, zornice, levý okraj pravého oka apod. Shluk několika těchto bodů představuje okraje úst, očí, tváře apod. Z těchto bodů lze následně např. vytvořit 3D model tváře, jenž může být vykreslen metodami počítačové grafiky. Predikované body pak zastávají funkci vertexů.

Výhodou zpracování tváře je většinou její jednoznačnost. Každý zdravý člověk má jeden pár očí a obočí, jedny ústa a jeden nos. Nezávisle na člověku se tyto orgány nachází relativně k sobě, v přibližně stejné vzdálenosti. Liší se zejména jen velikostí těchto orgánů. Komplikací mohou být úrazy tváře, které mohou měnit do jisté míry charakter tváře. Může jít o menší či větší rány, otoky (např. od zubů), teoreticky i plastika obličeje. Dále úspěšnost detekce obličejových bodů mohou komplikovat cizí předměty na tváři, jako jsou typicky brýle, nebo různé cvočky, piercing apod. V krajních případech může být kůže tváře potetovaná různorodými obrazci.

V současné době existuje přirozeně, i díky evoluci, množství technik a přístupů, vedoucí k detekci obličejových bodů. Každá metoda pracuje svým způsobem odlišně. Ve výsledku produkuje jiný počet obličejových bodů, s různou mírou přesnosti. Podle míry komplexity se rovněž liší v rychlosti evaluace vstupních dat. Nicméně všechny nějakým způsobem odkazují na předchozí kapitoly pojednávající o Počítačovém vidění a Strojovém učení.

V následujících podkapitolách se text odkazuje na aplikaci LANDMARK. Jde o vlastní aplikaci zastřešující zvolené metody predikce obličejových bodů. Její využití je myšleno zejména v rámci testování. Detailnější představení této aplikace je v kapitole 6.1.

V následujících podkapitolách budou představeny celkem tři metody detekce obličejových bodů. První je založena na úzkém propojení obličejových bodů s korespondujícími konturami obličeje. Druhá metoda je založena na regresních funkcích, respektive regresních stromech a predikuje obličejové body rychlostí v řádu milisekund. Třetí metoda je založena na nejmodernějších postupech detekce obličejových bodů a je schopna predikovat i 3D souřadnice obličejových bodů. První dvě zmiňované metody vycházejí z doporučené literatury, uvedené v zadání bakalářské práce. Poslední zmiňovaná metoda je zvolena autorem této práce.

5.1. Metoda založená na geometrii tváře

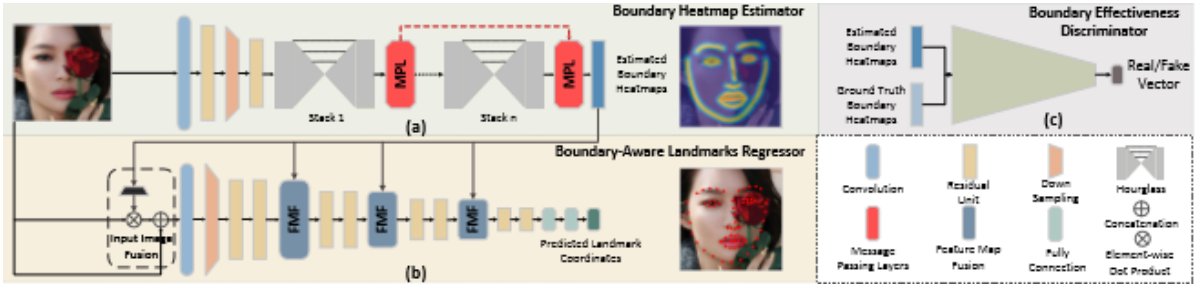
Autoři publikace [10] si kladli za cíl vyvinout efektivní algoritmus, schopný detekovat obličejové body nezávisle na typu mimiky nebo zakrytí tváře cizím objektem. Považují *„[…] hrany tváře jako geometrickou strukturu pomáhající predikci obličejových bodů na samém konci. Hrany velmi dobře popisují strukturu, která je konzistentní napříč obličeji a datovými množinami. Jsou také úzce spojeny s obličejovými body, protože většina obličejových bodů se nachází podél hraničních linek.“* Dle popsaných experimentů v publikaci [10] se ukázalo, že čím kvalitnější pravděpodobnostní mapy kontur byly, tím přesnější byly ve výsledku predikce obličejových bodů.

Výhodou je, že kontury jsou snadněji detekovatelné nezávisle na zakrytí části tváře nebo její rotaci. Z dostatečného množství obličejových bodů lze provést interpolaci kontur nezávisle na datových množinách, poskytujících typicky různé množství anotovaných obličejových bodů.

Metoda pracuje de facto ve dvou základních krocích, a to:

1. odhad/predikce kontur ze vstupního obrázku tváře,
2. predikce obličejových bodů s pomocí získaných kontur tváře.

Na vstupu konvoluční neuronové sítě (Obrázek 6) je snímek tváře částečně zakrytý růží. Podsíť tvořená generátorem kontur (Obrázek 6a), vygeneruje na základě vstupního snímku pravděpodobnostní mapu kontur tváře, úst, nosu, očí a obočí. Vstupní snímek a získané kontury jsou vstupními daty podsítě prediktoru obličejových bodů (Obrázek 6b), na jehož konci jsou finální souřadnice obličejových bodů.



Obrázek 6 – Architektura neuronové sítě. [10]

5.1.1. Generátor kontur

Generátor kontur (Boundary Heatmap Estimator) je část konvoluční neuronové sítě vytvářející pravděpodobnostní mapu kontur, a to jak samotné tváře, tak očí, obočí, nosu a úst. Tyto kontury představují geometrickou strukturu tváře a jejich kvalita je klíčová pro následující predikci obličejových bodů.

Základ architektury je sekvence podsítí typu Hourglass zakončená vrstvou – Message Passing Layer (MPL). Každá podsíť Hourglass odhaduje/predikuje jen jednu konturu např. spodní hranu oka atd. Kvalita kontur je ovšem negativně ovlivněna zejména při zakrytí tváře, proto je každá podsíť Hourglass zakončena vrstvou předávající informace (MPL). Význam této vrstvy spočívá v předávání informací ostatním vrstvám (MPL), čímž přispívají k lepší kvalitě a robustnosti výsledných kontur i při zakrytí tváře. Viditelné hrany tak mohou, na základě struktury tváře, pomoci ostatním podsítím typu Hourglass odhadnout neviditelné kontury. [10]

5.1.2. Kontradiktorní učení

Jestliže odhadnuté/predikované kontury nejsou přesné, dojde k negativnímu ovlivnění učení predikátoru obličejových bodů. Proto pro maximalizaci přesnosti odhadnutých kontur, respektive predikovaných obličejových bodů bylo využito tzv. kontradiktorního učení (adversarial learning).

Obrázek 6c znázorňuje klasifikační podsíť (Boundary Effectiveness Discriminator) rozhodující o přesnosti získaných kontur. Na vstupu jsou pravděpodobnostní mapy kontur (Estimated Boundary Heatmaps) a skutečné mapy kontur (Ground Truth Boundary Heatmaps). Cílem je, aby odhadnuté

kontury (výstup z generátoru kontur) byly tak přesvědčivé/kvalitní, že je klasifikátor označí za skutečné (Ground Truth Boundary Heatmaps). V nadsázce jde o „boj“, kdy se jedna podsíť snaží podvrhnutým vstupem zmást jinou podsíť (klasifikátor). Protože finální obličejové body kopírují linii kontur, snaží se klasifikátor predikovat, zda výsledné obličejové body budou co nejblíže skutečným konturám. [10]

5.1.3. Prediktor obličejových bodů

Učení prediktoru obličejových bodů využívá skutečné kontury jako vodítka pro učení příznaků. Skutečné obličejové body jsou interpolovány do hraničních linek tvořící binární mapu kontur. Interpolace je možná z dostatečného množství obličejových bodů, nezávisle na zvolené trénovací datové množině. Body ležící na hraniční lince jsou označeny jako 1, respektive 0 pokud jsou mimo linku. Na základě binární mapy kontur je vytvořena mapa vzdáleností jednotlivých pixelů k příslušné lince. Mapa vzdáleností je transformována do „skutečné“ mapy kontur pomocí Gaussovy funkce a standardní odchylky. Intenzita pixelů v mapě kontur je dána vzdáleností pixelu k příslušné kontuře.

Základní architektura sítě je tvořena čtyřmi propojenými podsítěmi 18-res. Na vstupu regresní sítě (Boundary-Aware Landmark Regressor) je spojení vstupního snímku tváře s odhadnutými konturami. Ke zlepšení výkonu sítě, dochází na počátku každé podsítě 18-res (FMF) k opětovnému spojení získaných kontur s aktuálními daty podsítě. [10]

5.1.4. Model

Výsledný model byl autory natrénován na datové množině WFLW, respektive její části pro trénování čítající 7 500 obličejů [11]. Ovšem model je možné pro srovnání úspěšně evaluovat i na typově jiných datových množinách obsahující jiná anotační schémata. Protože metoda je založená na využití kontur, dojde přirozeně ke sjednocení těchto schémat pouhou interpolací obličejových bodů do příslušných kontur.

Model pro svou evaluaci i případné dotrénování vyžaduje framework Caffe. Pracuje s jednokanálovým vstupem o rozlišení 256×256. Trénovací skript nebyl autory uvolněn. [10]

Pro potřeby testování v kapitole 7 je tento model označován jako WFLW.

5.1.5. Evaluační skript

Originální skript sloužil spíše k reprodukci výsledků publikace [10] na vlastním HW, a tak ho nebylo možné použít pro účely této práce. Z rozboru zdrojového kódu bylo také patrné, že algoritmus místo obvyklého detektoru tváře využívá manuálních anotacích k výpočtu ohraničení obličeje [12]. Dle GitHubu autorů [13] by měl být údajně ve vývoji skript umožňující evaluaci i na „cizí“ datové množině. Doposud (duben 2020) se však nestalo.

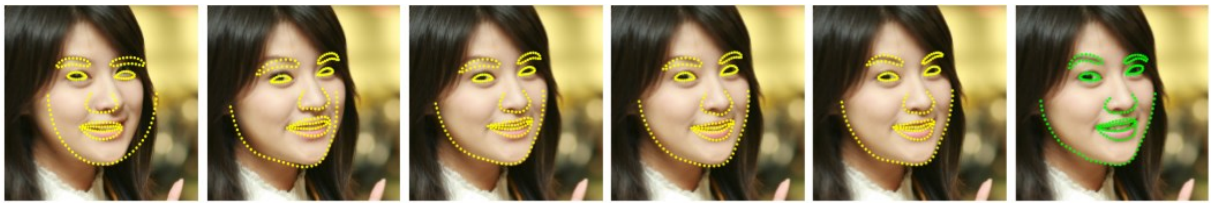
Výsledný evaluační skript je postaven na převzatém evaluačním skriptu [14], jenž umožňoval predikovat a následně vizualizovat obličejové body pro vstupní obrázek. Kroky predikce převzatého skriptu byly doplněny o detekci tváře. Proces predikce výsledného evaluačního skriptu začíná převodem vstupního obrázku do odstínů šedi, na základě detekce obličeje je snímek ořezán, normalizován a převeden na velikost 256×256. Získané predikce jsou uloženy ve formátu aplikace

LANDMARK, viz kapitola 6.1. Během zpracování každého obrázku jsou sbírány statistiky o rychlosti a úspěšnosti detekce obličeje.

Využívá se detektor tváře modulu dnn knihovny OpenCV založený na použití neuronových sítí, což umožňuje spolehlivější detekci, než např. detektory založené na HOG. V případě HOG detektoru by bylo zřejmě nutné řešit i situace, kdy hlava řidiče je od kamery odvrácená nebo najít detektor zvládající i detekce tváří z profilu. V opačném případě by došlo k selhání detekce tváře a tím i selhání predikce obličejových bodů. Pro zvýšení šancí na detekci, zejména malých obličejů, lze jako parametr skriptu předat souřadnice a velikost výřezu/oblasti původního obrazu. Předpokládá se tak, že řidič se na obrazu bude nacházet více či méně na stejném místě, čímž lze předem vyloučit pixely, na kterých se nepředpokládá výskyt obličeje.

5.2. Metoda založená na kaskádě regresních stromů

Metoda využívá k predikci obličejových bodů sadu regresních stromů, respektive kaskádu regresních funkcí učenou pomocí algoritmu Gradient Tree Boosting [15]. Výsledkem by měl být model schopný odhadnout souřadnice obličejových bodů rychlostí v řádu milisekund, s přesností blízkou současně nejmodernějším metodám v oblasti predikce obličejových bodů. Tato rychlost má být dosažena rozborem fundamentálních částí starších algoritmů a jejich efektivnějšímu začlenění do regresních funkcí. [16]



Obrázek 7 – Ukázka postupného zpřesňování obličejových bodů regresory kaskády (žlutě). Poslední obrázek znázorňuje Ground-truth (zeleně). [16]

5.2.1. Predikce a učení

Predikce obličejových bodů probíhá iterativně, viz Obrázek 7. Každý regresor v kaskádě zpřesňuje aktuální odhad pozic obličejových bodů. Snímek je na základě aktuálního odhadu tvaru tváře (souřadnic obličejových bodů) převeden do normalizovaného souřadného systému. Poté jsou extrahovány příznaky mající za úkol zpřesnit aktuální odhad obličejových bodů příslušným regresorem. Jako příznaky jsou považovány intenzity pixelů, jenž jsou vypočteny ze vstupního snímku a indexovány relativně k aktuálnímu odhadu obličejových bodů.

Každá regresní funkce je tvořená stromovým regresorem. V každém (nelistovém) uzlu stromu probíhá rozhodování na základě podmínky prahování intenzit dvou pixelů.

Učení první regresní funkce z kaskády je založena na trojici: vstupní obrázek, počáteční odhad obličejových bodů a velikost kroku aktualizace obličejových bodů. „*Počáteční struktura obličejových bodů může být jednoduše zvolena jako průměr vzatý z trénovacích dat, dále vycentrován a škálován podle ohraničení, jakožto výstup z obecného detektoru tváří.*“ Poté je tato trojice aktualizována a slouží jako trénovací data pro následující regresní funkci kaskády. Takto iterativním způsobem jsou trénovány všechny regresní funkce kaskády, dokud výsledné souřadnice obličejových bodů nejsou dostatečně přesné. [16]

5.2.2. Model

Dle publikace [16] byla jako trénovací datová množina použita sada HELEN. Trénování modelu na této množině, na jedno jádrovém CPU, zabralo cca. 1 hodinu. Evaluace jednoho obrázku trvala $\sim 1\ ms$. Výstupem je celkem 194 obličejových bodů. Experimenty autorů byly prováděny na datových množinách HELEN a LFPW. Naneštěstí nebyl autory uvolněn natrénovaný model, což vedlo k použití implementace metody knihovnou dlib.

Dle blogu knihovny Dlib [17] autor D. King uvádí, že:

- trénování modelu proběhlo na datové množině 300W [18] obsahující několik tisíc obrázků,
- hloubka kaskády stromů je 15,
- pomocí skriptu [19] je možné natrénovat vlastní model.

Přesnost a rychlost modelu má být srovnatelná s výsledky publikovanými samotnými autory metody v publikaci [16]. Výsledkem evaluace je celkem 68 obličejových bodů. Vzhledem k tomu, že model predikuje „pouze“ 68 obličejových bodů by se dalo očekávat, že rychlost evaluace by mohla být lepší než pouze srovnatelná.

V rámci testování bude na metodu, respektive tento model odkazováno označením dlib.

5.2.3. Evaluační skript

Příklad evaluačního skriptu knihovny Dlib se nachází na webu [20]. Aby došlo k unifikaci skriptů, jakožto i korektní spolupráci s aplikací LANDMARK (Kapitola 6.1), byla část skriptu týkající se predikce obličejových bodů, zakomponována do struktury kódu předchozí metody. Respektive došlo k nahrazení prediktoru obličejových bodů prediktorem této metody.

Logika samotného detektoru obličeje zůstala z předchozí metody nepozměněna. Dle blogu [17] byl při trénování použit detektor tváří knihovny Dlib. Dlib nabízí detektor tváří, detekující tváře z čelního pohledu, na bázi HOG a SVM, který nemusí fungovat zejména při natočení tváře řidiče. Proto aby se minimalizovaly neúspěšné detekce, byl ponechán detektor předchozí metody (Kapitola 5.1).

5.3. Metoda predikující 2D a 3D obličejové body

Poslední metoda přináší možnost predikovat jak 2D, tak i 3D obličejové body. Model neuronové sítě vzešel z publikace [8] pojednávající o výkonu velmi hlubokých neuronových sítích napříč datovými množinami. Jako jeden z cílů této publikace bylo vytvoření a natrénovaní výkonného modelu pro detekci obličejových bodů.

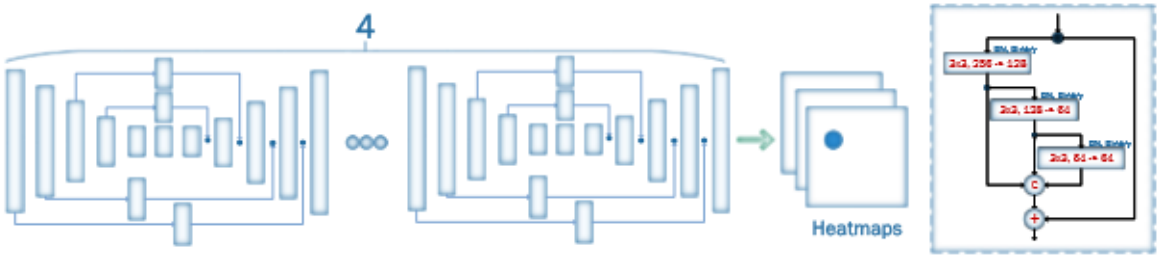
Pro účely této bakalářské práce bude testování zaměřeno pouze na 2D obličejové body, a to z důvodu ostatních testovaných metod, které predikují pouze 2D souřadnice.

5.3.1. 2D a 3D síť

Dle publikace [8] se autoři pokusili vytvořit jednu z nejvýkonnějších neuronových sítí pro detekci obličejových bodů. Autoři ji pojmenovali jako Face Alignment Network (FAN), viz Obrázek 8. Je založena na jedné z nejmodernějších architektur neuronových sítí, navržené pro Human Pose Estimation.

Je tvořena sérií neuronových podsítí typu Hourglass. V sérii se tento typ neuronové podsítě nachází celkem čtyřikrát. Struktura konvolučních bloků je hierarchická, paralelní a víceúrovňová.

Použitím výše zmíněné architektury vzešly celkem dva modely založené na konvolučních neuronových sítích, na jejichž výstupech je celkem 68 2D, respektive 3D souřadnic obličejových bodů. [8]



Obrázek 8 – Model FAN (vlevo). Hierarchická struktura konvolučního bloku (vpravo). [8]

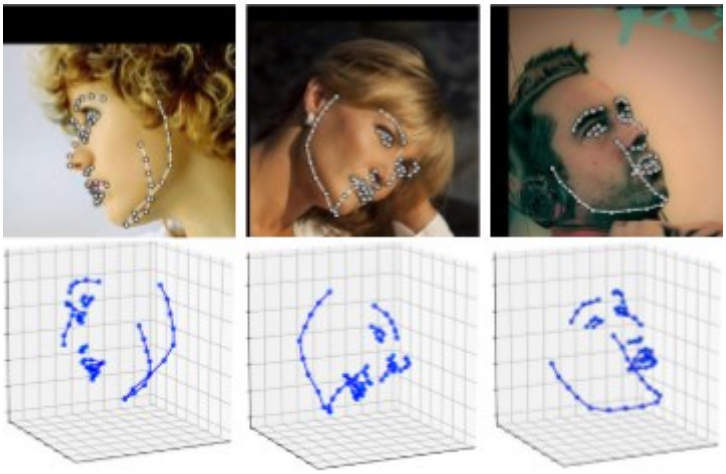
5.3.2. Z 2D do 3D

Za účelem natrénování neuronové sítě pro predikci obličejových bodů ve 3D, byla vytvořena další neuronová síť – 2D-to-3D-FAN. Jde de facto o konvertor souřadnic vstupních 2D obličejových bodů do 3D. Respektive jde o získání 2D projekcí 3D souřadnic. Tyto 2D projekce souřadnic jsou použity k vytvoření nové datové množiny – LS3D-W. Mělo by jít o největší datovou množinu vztahující se k problematice 3D obličejových bodů, čítající cca. 230 000 obrázků, získanou sjednocením všech tehdejších (2017) datových množin. [8]

Publikace [8] také uvádí rozšíření 2D-to-3D-FAN o schopnost predikovat také *z* souřadnici (hloubku) obličejových bodů. Výstupem už tím pádem není „pouhá“ 2D projekce 3D souřadnici, nýbrž plnohodnotné 3D souřadnice.

Toto rozšíření, z 2D-to-3D-FAN do Full-2D-to-3D-FAN, spočívá v přidání podsítě založené na architektuře ResNet-152, která má za úkol predikovat, již zmíněnou souřadnici hloubky. Vstupem do této sítě jsou pravděpodobnostní mapy, jakožto výsledek sítě 2D-to-3D-FAN, a také RGB obrázek obličeje. Tato podsíť je při predikci hloubky vedena právě získanými pravděpodobnostními mapami, které určují, na kterých pozicích by hloubka měla být predikována. [8]

Ukázku predikovaných 3D souřadnic z 2D snímků znázorňuje Obrázek 9.



Obrázek 9- Vizualizace 3D souřadnic obličejových bodů z 2D snímku. [8]

5.3.3. Model

Pro natrénování modelů 2D-FAN a 3D-FAN byly využity datové množiny 300W-LP-2D, respektive 300W-LP-3D pro 3D-FAN. Pro dotrénování 2D-FAN byla rovněž využita datová množina 300-W. Konvoluční neuronová síť 2D-to-3D-FAN byla natrénována pomocí datové množiny 300-W-LP obsahující jak 2D, tak i 2D projekce 3D anotací pro daný obrázek. [8]

Pro účely testování se na tento model odkazuje označením 2D_FAN.

5.3.4. Evaluační skript

Evaluace je v tomto případě řešena skriptem v Pythonu. Což z hlediska rychlosti, může být nevýhoda oproti předchozím metodám (Kapitola 5.1 a Kapitola 5.2) mající skripty v C++. Metoda nabízí i přidružený detektor tváře, založený na konvolučních neuronových sítích.

Skript rovněž implementuje výstup souřadnic obličejových bodů ve formátu aplikace LANDMARK (Kapitola 6.1). Taktéž sbírá statistiky o rychlosti predikce obličejových bodů.

6. Testovací prostředí

Testovací prostředí je tvořeno vyvinutou desktopovou aplikací LANDMARK (Kapitola 6.1), skripty predikující obličejové body, skriptem zpracovávající statistiky predikcí a skriptem pro zadávání manuálních anotací. Tento testovací řetězec a jeho fáze jsou detailněji popsány v příslušné podkapitole.

Detailnějšímu rozboru nastíněných témat je věnováno několik následujících podkapitol.

6.1. Aplikace

Za účelem procházení pořízených videosekvencí a vizualizaci predikovaných obličejových bodů byla vyvinuta desktopová aplikace LANDMARK.

Aplikace je z důvodu přenositelnosti, a také z důvodu udržení rozumné náročnosti vývoje, vyvinutá v jazyce Java. Jako GUI framework je použit JavaFX. Základní architektonický model aplikace je MVC.

Hlavním smyslem využití této aplikace je především empirické testování predikovaných obličejových bodů na pořízené datové množině. Zabraňuje rovněž duplikaci logiky vizualizace mezi jednotlivé skripty příslušných metod, predikující obličejové body. V neposlední řadě aplikace představuje jakýsi centrální bod řešené problematiky.

Výhodou této aplikace je rovněž oddělení od samotné technologie skriptů predikující obličejové body. Skripty tak mohou být vyvinuty v libovolném programovacím či skriptovacím jazyce. Pokud skripty dodrží formát svých výstupních souborů, lze tyto soubory zpracovat v aplikaci.

Následující podkapitoly se zběžně věnují funkcionalitě aplikace.

6.1.1. Práce se vstupy

Ačkoliv je zamýšleno testovat metody detekce obličejových bodů na videosekvencích, aplikace pracuje s jednotlivými snímky původní videosekvence. Tento způsob přináší výhody ve formě možnosti načíst přímo kýžený snímek např. za účelem ladění. Od uživatele se nicméně vyžaduje extrahování snímků z pořízených videosekvencí do formátů typu JPEG, PNG nebo BMP.

Aby uživatel nepřišel o možnost prohlédnout si celou sadu snímků původní videosekvence, nabízí aplikace rovněž možnost pracovat s celou množinou zvolených snímků. Relativní cesty zvolených snímků jsou uloženy v souborech sekvence s příponou sequence. Vždy se ukládají jen relativní cesty k místu uložení souboru sekvence. V době běhu aplikace jsou načteny pouze relativní cesty ke snímkům. Samotný snímek je načten až ve chvíli, kdy má být zobrazen uživateli.

Přestože jde jen o textové soubory se seznamem relativních cest, aplikace umožňuje tyto soubory pohodlně z grafického rozhraní vytvářet a editovat.

Využití takto spravovaných sekvencí snímků je i mimo aplikaci LANDMARK, a sice v případě predikování obličejových bodů evaluačními skripty. Jde tedy o vstupní soubor s relativními cestami snímků, pro které je zamýšleno predikovat obličejové body.

6.1.2. Správa metod predikce

Aplikace spravuje seznam dostupných metod, jejich základní atributy a nastavení. Je to zejména název metody, jména autorů, adresa webu a základní popis. Nastavení se týká samotné vizualizace, viz kapitola 6.1.3. Uživatel vybírá metody z tohoto seznamu při načítání obrazových dat. Na základě této volby aplikace načte příslušné soubory s predikovanými obličejovými body.

Jako permanentní úložiště slouží XML soubor nacházející se v adresáři aplikačních dat aplikace. Ve Windows je to adresář …/<uživatel>/AppData/Roaming/LANDMARK. V případě Linuxu je to adresář /home/<uživatel>/.config/LANDMARK.

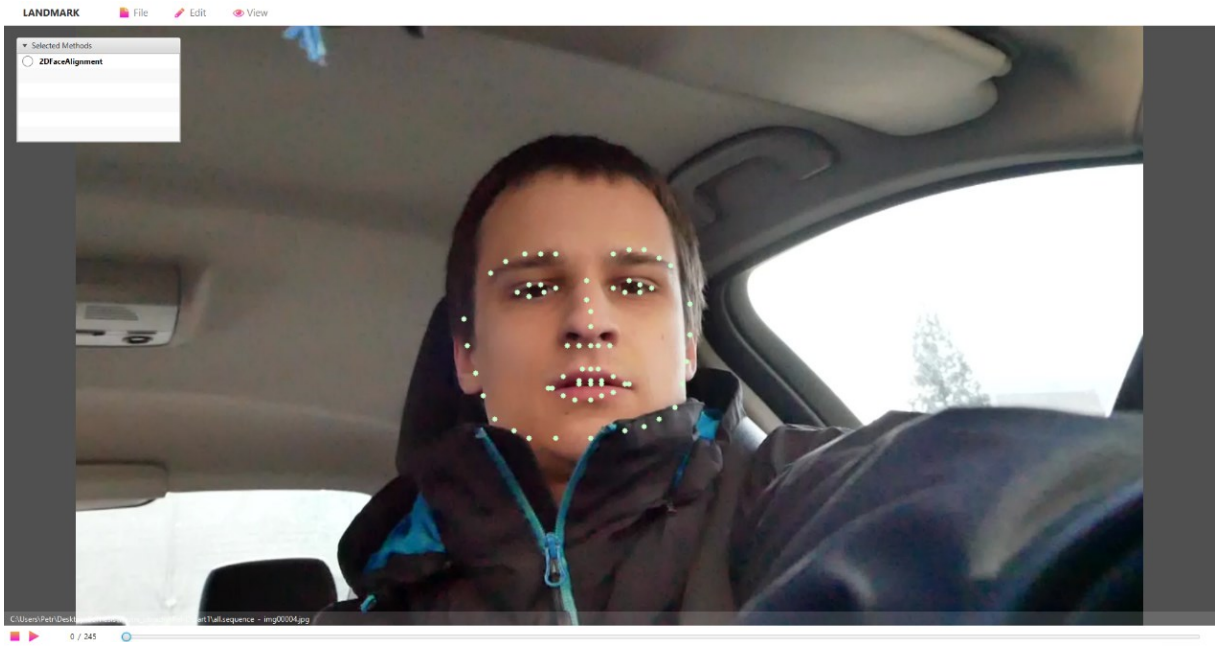
6.1.3. Vizualizace

V hlavním okně aplikace se nachází přehrávač. Tento přehrávač, mimo přehrání zvolené obrazové sekvence, dokáže zastat i funkci prohlížeče právě v případě testování pouze jednoho zvoleného obrázku.

Aplikace obsahuje rovněž zabudovaný renderer obličejových bodů, viz Obrázek 10. Tyto body jsou vizualizovány skutečnými body a jsou parametrizovány uživatelským nastavením. Nastavení se omezuje na volbu barvy a vykreslované velikosti bodu. Rovněž je možné vykreslování množiny obličejových bodů omezit pomocí masky na zvolené body. Nastavení je specifické pro každou metodu.

Mimo samotné obličejové body lze rovněž vykreslit jejich příslušné indexy nebo detekované ohraničení obličeje.

Pro lepší orientaci uživatele během přehrávání obrazové sekvence, je možné v aplikaci zobrazit legendu. Legenda se sestává ze seznamu zobrazovaných metod, respektive názvu metody a zvolené barvy vykreslovaných obličejových bodů.



Obrázek 10 – Hlavní okno aplikace LANDMARK s přehrávačem sekvencí. Jsou zde znázorněny predikované body (zelené tečky).

6.2. Zpracování manuálních anotací

Představené metody detekce obličejových bodů jsou porovnávány vůči ručně zadaným obličejovým bodům, respektive jejich souřadnicím. Musí tedy existovat nástroj, pomocí něhož bude možné požadované obličejové body zadat a vhodně uložit pro další zpracování.

Na internetu jsou dostupné jak placené, tak neplacené nástroje pro ruční anotace obrázků. Avšak po rešerši dostupných nástrojů nebyl zvolen ani jeden, a bylo tak přistoupeno k vývoji vlastního nástroje, respektive konzolového skriptu. Výhodou této volby je mnohem snazší integrace do ekosystému testovacího prostředí.

Ačkoliv by se pro účel zadání souřadnic obličejových bodů hodil grafický nástroj, byl v celku rychle vyvinut konzolový skript v jazyce Python. Skript manual_annotation.py umožňuje:

- manuální zadání souřadnic obličejových bodů,
- asistované manuální zadání souřadnic obličejových bodů,
- vytvoření mapování mezi indexy manuálně zadaných obličejových bodů a indexy predikovaných obličejových bodů,
- provedení porovnání manuálně zadaných souřadnic obličejových bodů s korespondujícími predikovanými body.

Ruční vkládání souřadnic obličejových bodů je podporováno hned dvěma způsoby. Liší se zejména tím, že asistované zadávání příslušných souřadnic uživateli nabízí k příslušnému bodu již predikované souřadnice zvolenou metodou. Využívá se tak již predikovaných souřadnic, které jsou manuálně zpřesněny. Protože se souřadnice zadávají textově do terminálu, je nutné tyto souřadnice získat jiným způsobem. Obecně lze využít libovolnou grafickou aplikaci, umožňující získat 2D souřadnice pixelu. Zadané souřadnice se následně ukládají v totožném formátu, v jakém jsou uloženy predikované souřadnice. Je tedy možné i tyto ručně zadané obličejové body vizualizovat aplikací LANDMARK.

Jelikož se liší počty obličejových bodů jak mezi zvolenými metodami predikce, tak ručně zadanými obličejovými body. Bylo nutné vytvořit mapování, a to mapování indexů představující obličejové body metod na příslušné indexy ručně zadaných bodů. Toto mapování je nedílnou součástí následného porovnání souřadnic.

Vždy se porovnávají predikované souřadnice obličejových bodů vůči korespondujícím ručně zadaným bodům. Aby porovnání bylo možné, využívá se již zmíněné mapování.

6.3. Zpracování statistik

Evaluační skripty příslušných metod po svém skončení ukládají do souboru statistiky. V tomto případě pro účely testování jde o rychlost predikce obličejových bodů. Ve výsledku se jedná o soubor obsahující data ve formátu klíč-hodnota, kde klíčem je relativní cesta ke snímku a hodnotou je čas predikce. Pro každý zpracovaný snímek je uložena rychlost predikce v sekundách nebo případně milisekundách.

Pokud by se do testování zahrnula celá sada snímků, pro kterou existuje statistika, mohl by přímo tento soubor obsahovat i aritmetický průměr časů predikce. Ve skutečnosti, jak bude blíže specifikováno v kapitole o testování, se sady snímků dále vzorkují. Je tedy potřeba disponovat

skriptem, který z několika zvolených souborů statistik, dokáže vyextrahovat příslušné časy predikce vybraných vzorků a vypočítat (průměrný) čas predikcí.

Ručně je ovšem nutné tyto soubory pročistit, protože pro některé snímky mohl detektor obličeje detekovat více tváří. V souboru se tedy může nacházet více časů predikce pro jednotlivé vzorky. Je tedy nutné ponechat maximálně jen jednu časovou hodnotu pro vzorek.

Uvedená funkcionalita je představována Python skriptem stats.py.

6.4. Testovací řetězec

Z předcházejících kapitol je zřejmé, že samotná procedura testování není zrovna triviální. Celkově je tvořena mnoha na sebe navazujícími kroky. Proto pro shrnutí a lepší představu o procesu testování je zde v krocích uveden algoritmus znázorňující proces testování.

1. **Predikování**
   1.1. Pomocí aplikace LANDMARK dojde k sestavení požadované množiny snímků, pro kterou se budou predikovat souřadnice obličejových bodů. Výsledkem je soubor *.sequence.
   1.2. Následuje spuštění příslušného evaluačního skriptu metody pro sestavenou množinu snímků v kroku 1.1. Výsledkem jsou soubory se souřadnicemi obličejových bodů, soubory s detekovanými ohraničeními obličeje a jeden souhrnný soubor se statistikami.
2. **Anotování**
   2.1. Pro sestavenou množinu snímků jsou vytvořeny manuální anotace obličejových bodů. (Skript manual_annotation.py)
   2.2. Pro testované modely jsou vytvořena mapování mezi schématy obličejových bodů modelů a schématem ručních anotací. (Skript manual_annotation.py)
3. **Vyhodnocování**
   3.1. Porovnání predikovaných souřadnic obličejových bodů vůči ručně anotovaným souřadnicím. (Skript manual_annotation.py)
   3.2. Zjištění průměrných času predikce. (Skript stats.py)

7. Testování

V kapitole o testování bude čtenář nejprve seznámen s metodikou testování, respektive způsobem vzorkování datové množiny a se způsobem výpočtu a vyjádření chyby predikcí.

Následovat bude podkapitola věnující se vizuálnímu rozboru některých testovaných situací. V závěru kapitoly budou uvedeny celkové výsledky zvolených metod predikce obličejových bodů. Testování je zaměřeno nejen na přesnost predikovaných obličejových bodů, ale také na rychlost.

Testování probíhalo na sestavě PC:

- CPU Intel i5-4460 3,20 GHz – 4 jádra,
- 8 GB RAM.

7.1. Metodika testování

Predikované obličejové body, respektive souřadnice těchto bodů, jsou porovnány s ručně zadanými obličejovými body představující referenční hodnoty (anotace). Míra, jakou se liší predikované a ručně zadané souřadnice, určuje vypočtená chyba predikce.

V následujících podkapitolách je popsán způsob vzorkování dat, jaké obličejové body jsou testovány a jakým algoritmem je stanovena chyba predikcí.

7.1.1. Vzorkování

Primárním zdrojem vzorků jsou snímky pořízené v reálných podmínkách automobilu. Snímky z laboratorních podmínek byly pro svou variabilitu mimiky využity jako sekundární zdroj. Za účelem redukce a zároveň zachování reprezentativního vzorku dat, byla pro účely testování datová množina zredukována, a to několikrát.

Datová množina zachycená v reálných podmínkách byla zredukována již při extrahování snímků z videa. Extrahoval se každý třetí snímek. Tím došlo k prvotní minimalizaci velmi podobných snímků jdoucích v sekvenci.

Další vzorkování je provedeno až po získání predikcí obličejových bodů. Je založeno na pozorování, načež jsou selektivně zvoleny unikátní snímky. Unikátní ve smyslu pózy, mimiky, nasvícení, zakrytí tváře cizím objektem apod. Zaměření bylo kladeno zejména na extrémy v těchto situacích ve smyslu natočení hlavy nebo ostrých stínů v obrazu. Bylo vybráno 120 vzorků z reálných podmínek automobilu a 17 doplňujících vzorků z prostředí laboratoře. Tyto selektivně zvolené vzorky jsou následně ručně anotovány.

Dávalo by smysl selektovat data již před samotnou predikcí obličejových bodů a celý proces predikce by tím byl značně urychlen. Problémem však je, že predikce následuje až jako další krok po úspěšné detekci tváře. Snímky s neúspěšnou detekcí tváře jsou z testování implicitně vyřazeny, neboť je testování zaměřeno pouze na obličejové body. Podmínkou testování je, že snímek se testuje jen ve chvíli, kdy jsou pro daný snímek dostupné predikce obličejových bodů testovaných metod. Splnění této podmínky je tedy přímo závislé na úspěšnosti obou detektorů. Zejména videosekvence pořízená v noci je zatížena velkou neúspěšností detektorů. V takovém případě je tedy i značně obtížné náhodně

zvolit vzorek před predikcemi, respektive detekcemi tváří. S velkou pravděpodobností by pro takový náhodný vzorek nezafungoval ani jeden z detektorů a snímek by byl ve výsledku bez predikcí.

Jednoznačně největší nevýhodou je samotný čas strávený predikováním. Naštěstí je datová množina relativně kompaktní a lze získat v rozumném čase predikce zvolených metod. Výhodou ovšem je, že mohou být do testování zahrnuty snímky představující určitou výzvu pro prediktor obličejových bodů. Vzorkování tedy není náhodné, ale záměrně selektivní.

7.1.2. Manuální anotace

Manuální anotace slouží jako reference k porovnání predikovaných obličejových bodů. Nejen protože je manuální anotace poměrně pracná, ale také kvůli různým sadám obličejových bodů predikovaných zvolenými metodami, byly vybrány jen některé společné body, viz Obrázek 11. Jde o:

- koutky očí,
- špičku nosu,
- kraje obočí,
- koutky úst,
- špičku brady.

Obrázek 11 – Manuální anotace na jednom z testovaných snímků v reálných podmínkách automobilu.

7.1.3. Měření přesnosti

Přesnost predikce je v této práci vyjádřena jako vzdálenost mezi predikovaným a anotovaným obličejovým bodem. Základem výpočtu je Pythagorova věta.

$$Průměrná\ vzdálenost\ (chyba) = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left(P_{x,i} - M_{x,i}\right)^2 + \left(P_{y,i} - M_{y,i}\right)^2} \quad (1)$$

Celkově jde o vyjádření průměrné vzdálenosti pro jeden typ obličejového bodu (např. špička nosu), pro N testovaných snímků, respektive N tváří, kde P je predikovaný bod a M korespondující manuálně anotovaný bod. Tímto způsobem je získáno celkem 12 vzdáleností, zvlášť pro každý typ testovaného obličejového bodu.

7.2. Pozorování

V rámci pozorování jsou rozebrány vybrané situace z reálných podmínek automobilu. Do pozorování jsou zahrnuty zejména ukázky vzorků, jenž představují pro metody predikce obličejových bodů výzvu.

Ukázky vzorků nejsou nutně reprezentativní pro rozebíraný problém ve smyslu uvedených chyb predikcí. Z testové množiny vzorků byly vybrány snímky vhodné pro ilustraci problémů příslušných podkapitol. Chyby predikcí pro ukázkové vzorky mohou představovat odchylku, kdy použitím jiného snímku s podobnou pózou může být chyba predikce větší nebo menší. V kapitole 7.3 jsou v rámci shrnutí výsledků uvedeny průměrné chyby modelů pro celou sadu testových vzorků, čímž dojde k minimalizaci případných odchylek chyb predikce.

7.2.1. Variabilita póz

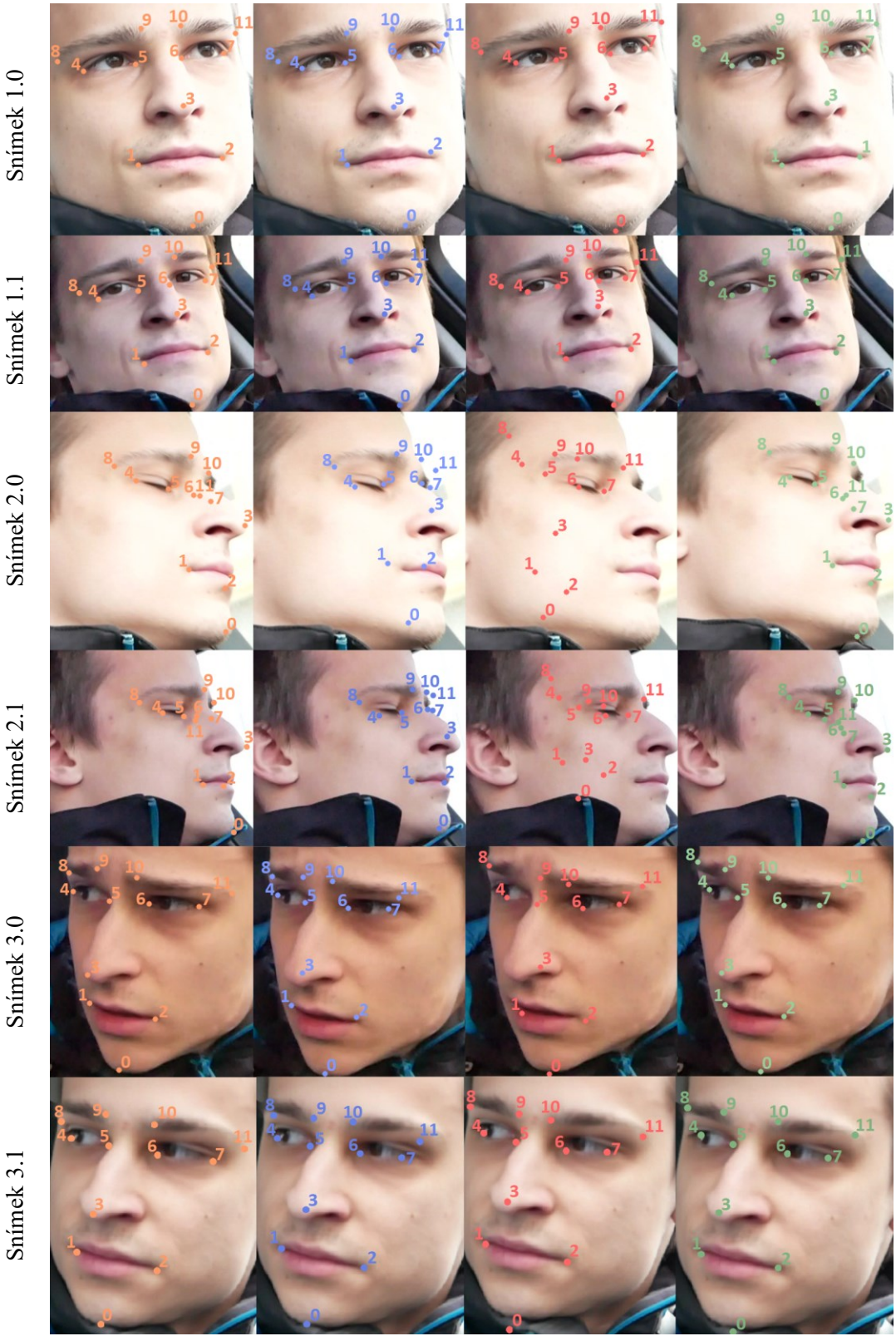
Vytvořená datová množina z reálných podmínek automobilu obsahuje poměrně velké množství póz. Pózou se myslí jak mimika, tak různá natočení obličeje ke snímači. Videosekvence byly pořízeny ve městě, kde se řidič často rozhlíží na křižovatkách. Vybrané vzorky tedy nebyly ochuzeny i o poměrně extrémní případy, kdy je obličej ke snímači v úhlu $\sim 90°$.

Obrázek 12 představuje celkem tři unikátní pózy, pro které jsou predikovány obličejové body. Každá póza je zastoupena dvěma různými snímky. Snahou je, aby na každé dvojici byla velmi podobná póza, čímž dojde k minimalizaci případných odchylek. Kritérium výběru těchto tří póz bylo ukázat tvář z různých úhlů za účelem zjištění limitů zvolených metod predikce obličejových bodů. Podrobnější výsledky testů nabízí Tabulka 1.

Snímky 1.0 a 1.1 (Obrázek 12) obsahuje tvář téměř z čelního pohledu. Predikce všech tří testovaných modelů mají průměrnou chybu do $8{,}0\ px$. Lze tedy usoudit, že téměř čelní pohled nepředstavuje pro zvolené metody predikce výzvu.

Snímky 2.0 a 2.1 (Obrázek 12) zobrazuje tvář svírající se snímačem úhel $\sim 80°$. Z testovaných modelů je nejblíže manuálním anotacím model 2D_FAN s chybou $10{,}7\ px$. V případě tohoto modelu je vychýlen bod 0. Body oka (6 a 7) jsou ovšem na odvrácené straně tváře pod větším sklonem, než by odpovídalo celkovému natočení tváře. Selháním je model dlib, jehož predikce jsou značně vychýlené od manuálních anotací. Model WFLW vykazuje chyby zejména pro predikce bodů 0–3, 6–7 a 10–11.

Obličej na Snímcích 3.0 a 3.1 (Obrázek 12) je v ostrém úhlu do 30°. Sémanticky jsou víceméně správně predikce modelů WFLW a 2D_FAN. Významnější odchylky lze vidět u obličejových bodů 0–5 a 9 modelu dlib.



Obrázek 12 – Ukázka póz z reálných podmínek automobilu.
(a) – Manuální anotace; (b) – model WFLW; (c) – model dlib; (d) – model 2D_FAN

Tabulka 1 – Výsledky testování pro Obrázek 12 (Červeně chyba ≥ 20 px).
Uvedené chyby jsou průměrem dvou snímků, specifikovaných v záhlaví sekce.

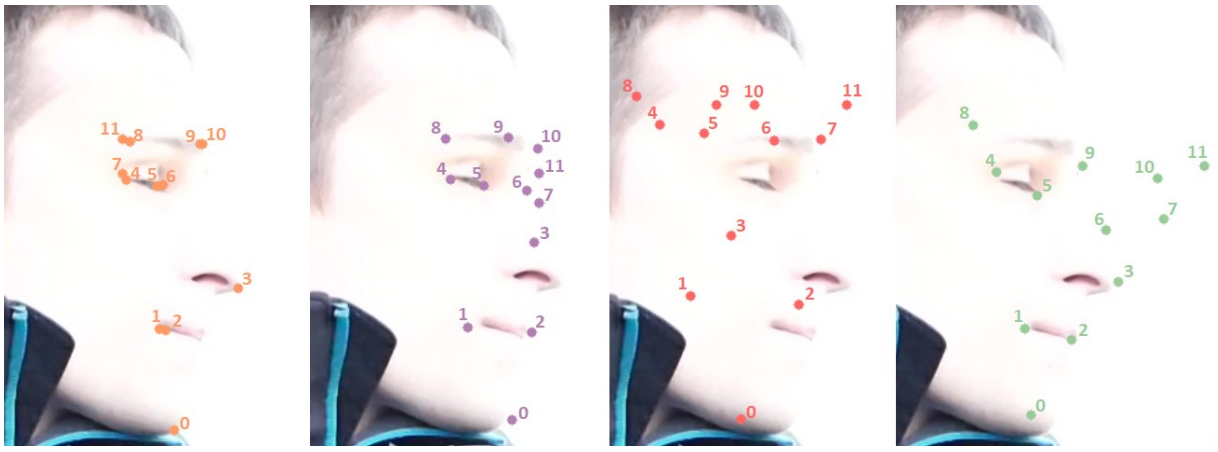
Model	Chyby predikce obličejových bodů [px]												
	0	1	2	3	4	5	6	7	8	9	10	11	0–11
Snímky 1.0 a 1.1													
WFLW	2,1	3,9	7,2	2,2	9,8	3,7	10,5	3,7	12,4	9,2	3,9	1,8	5,9
dlib	9,1	3,6	2,5	6,2	13,2	4,3	9,1	3,4	6,5	8,5	10,4	7,7	7,0
2D_FAN	5,8	2,7	4,2	1,8	10,0	4,9	12,3	6,5	15,8	11,0	8,9	11,6	8,0
Snímky 2.0 a 2.1													
WFLW	30,9	16,7	23,0	35,3	9,3	6,3	25,4	17,7	8,5	11,3	17,2	48,5	20,8
dlib	151,7	116,8	103,8	156,3	63,2	68,6	50,8	39,6	57,6	84,6	77,6	40,2	84,2
2D_FAN	25,4	4,4	9,2	4,7	7,4	3,2	15,2	22,9	14,9	10,7	5,8	4,5	10,7
Snímky 3.0 a 3.1													
WFLW	3,8	3,6	5,8	16,6	2,7	8,4	5,1	23,3	7,9	9,0	10,4	51,0	12,3
dlib	19,1	13,1	8,8	29,6	13,0	8,9	12,1	13,0	11,6	23,7	20,5	17,9	16,0
2D_FAN	14,7	4,9	7,8	4,8	8,6	6,1	6,8	17,9	20,3	9,9	5,0	33,2	11,7

Z pozorování této podkapitoly vyplývá, že všechny testované metody vykazují v průměru jen malé chyby v predikcích pro tváře blízké čelnímu pohledu. Model dlib by mohl odhadem relativně přesně predikovat obličejové body pro tváře svírající úhel se snímačem do ~30°. Predikce modelu WFLW vykazují významnou chybu pro Snímky 2.0 a 2.1, jeho limity by mohly být pro tváře v úhlu se snímačem do ~65°. Velmi přesné predikce prokázal model 2D_FAN.

7.2.2. Přeexponování

Chyby v expozici snímku přinášejí problémy v podobě ztráty detailu. Přeexponování snímku znamená, že obraz obsahuje velké plochy jednolitých bílých pixelů bez textury.

Pro představu, situaci znázorňuje Obrázek 13. Kvůli ostrému zimnímu slunci pronikající skrze boční okno, dochází ke značnému osvětlení tváře řidiče. Důsledkem tohoto nasvícení je zcela úplný zánik detailů a kontur tváře. Samotná tvář je navíc zachycena v úhlu ~90°. Souřadnice obličejových bodů viditelné části tváře se tak částečně kryjí se souřadnicemi odvrácené části tváře, viz Obrázek 13a.



(a) (b) (c) (d)

Obrázek 13 – Přeexponovaná tvář.
(a) – Manuální anotace; (b) – model WFLW; (c) – model dlib; (d) – model 2D_FAN

Z pozorování je patrné, že nejblíže manuálním anotacím je model WFLW (Obrázek 13b), s průměrnou chybou 28,6 px. Ovšem i v tomto případě má model potíže se zakrytými obličejovými body (body 2–3 a 6–11). Ve výsledku horší predikce nabízí model 2D_FAN, s průměrnou chybou 48,4 px. Zakryté body se v tomto případě ani sémanticky neblíží své správné lokaci. Model dlib z takového profilu tváře nepredikoval správně ani jeden bod. Detailnější přehled chyb obsahuje Tabulka 2.

Tabulka 2 – Výsledky testování pro Obrázek 13 (Červeně chyba ≥ 20px).

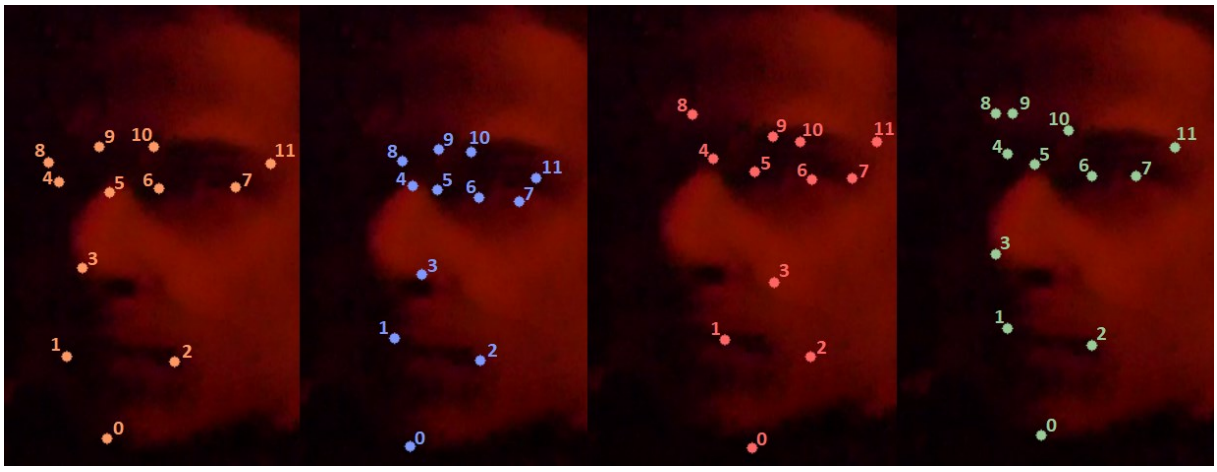
Model	Chyby predikce obličejových bodů [px]												
	0	1	2	3	4	5	6	7	8	9	10	11	0–11
WFLW	14,1	10,2	**36,5**	**40,4**	3,6	5,0	**34,8**	**81,6**	4,9	13,1	15,4	**83,1**	**28,6**
dlib	**36,9**	**70,0**	**26,9**	**104,8**	**76,8**	**66,8**	**35,0**	**76,0**	**92,3**	**84,9**	**55,5**	**95,9**	**68,5**
2D_FAN	17,7	5,1	**28,9**	9,2	7,1	9,9	**68,3**	**142,6**	**26,9**	18,4	**75,0**	**172,3**	**48,4**

7.2.3. Podexponování

Jde opět o chybnou expozici, ovšem opačnou k přeexponování. Podexponování znamená, že obrázek obsahuje mnoho tmavých míst, postrádající detail, viz Obrázek 14

Řešením může být delší doba, po kterou je snímač vystaven světlu. Výsledkem bude méně tmavých míst v obraze. Naneštěstí zpracování snímků v reálném čase vyžaduje pořízení snímků v co nejkratším čase. Dále je nutné po dobu snímání zajistit maximálně statický obraz, jinak dojde k rozmazání snímku.

Dalším řešením je zvýšení hodnoty ISO snímače, čímž současně dojde ke zvýšení šumu. K tomuto řešení bylo přistoupeno i v případě pořízení videosekvence v noci, viz Obrázek 14.



(a) (b) (c) (d)

Obrázek 14 – Podexponovaná tvář.
(a) – Manuální anotace; (b) – model WFLW; (c) – model dlib; (d) – model 2D_FAN

Nejhůře predikoval obličejové body model dlib (Obrázek 14c), jeho průměrná chyba je 46,3 px. Nejvíce se správné pozici vzdaluje bod 3 s chybou 81,7 px. Za druhý nejhorší výsledek lze považovat průměrnou chybu 25,2 px modelu WFLW. V tomto případě se nejvíce vychýlil bod 8, a to o 44,2 px.

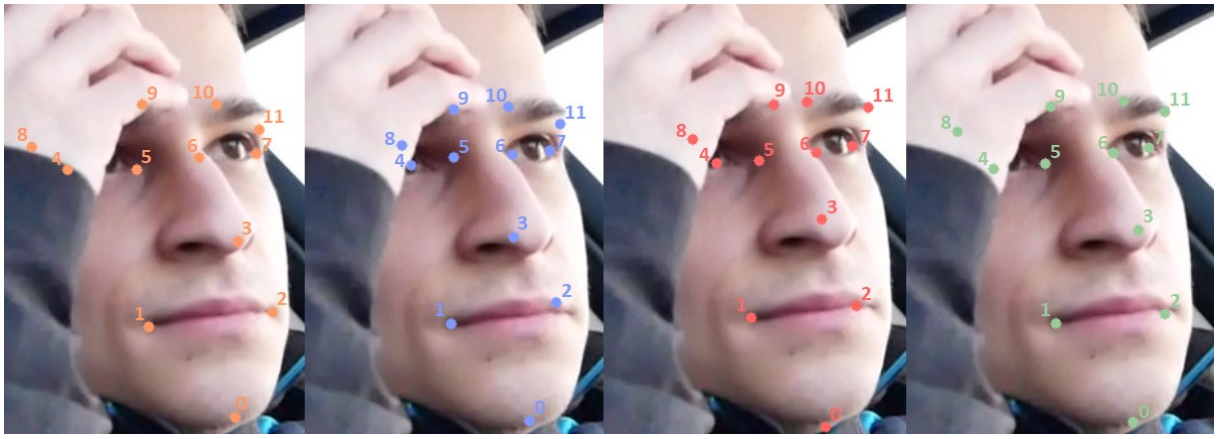
Nejlépe dopadly predikce modelu 2D_FAN s průměrnou chybou 15,6 px. Ovšem predikce bodů 1,4 a 8, tedy body na odvrácené straně tváře, jsou predikovány se značnou chybou, viz Tabulka 3.

Tabulka 3 – Výsledky testování pro Obrázek 14 (Červeně chyba ≥ 20px).

Model	Chyby predikce obličejových bodů [px]												
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**$\overline{0\text{–}11}$**
WFLW	8,2	**28,1**	7,3	**33,4**	**44,1**	**24,9**	**20,2**	14,8	**44,2**	**33,3**	16,8	**26,4**	**25,2**
dlib	**46,2**	**53,1**	**36,5**	**81,7**	**50,6**	**43,4**	**49,1**	**21,2**	**49,9**	**65,0**	**44,4**	14,8	**46,3**
2D_FAN	13,9	**22,8**	5,0	4,2	**27,8**	15,2	12,2	13,0	**37,8**	18,2	5,4	11,2	15,6

7.2.4. Částečné zakrytí tváře

Zakrytí tváře nelze zejména v automobilu ignorovat, neboť za slunečních dní řidiči často řídí vozidlo se slunečními brýlemi. Dalším způsobem zakrytí tváře může být např. ruka nebo límec bundy, viz Obrázek 15.



(a) (b) (c) (d)

Obrázek 15 – Částečné zakrytí tváře rukou.
(a) – Manuální anotace; (b) – model WFLW; (c) – model dlib; (d) – model 2D_FAN

Lze vypozorovat, že průměrně nejpřesnější predikce obličejových bodů, s chybou 8,7 px, poskytuje model 2D_FAN, viz Obrázek 15d. Největší chyby predikcí modelu 2D_FAN jsou pro body 4, 8 a 11. Modely WFLW a dlib chybovaly zejména v predikcích zakrytých bodů (4,8 a 9).

Lze si také povšimnout, že predikce obličejového bodu 6 testovaných modelů, nejsou zcela na správném místě. Bod 6 je zakrytý nosem, tudíž by jeho predikce neměly být na viditelném místě. Kompletní vyčíslení chyb obsahuje Tabulka 4.

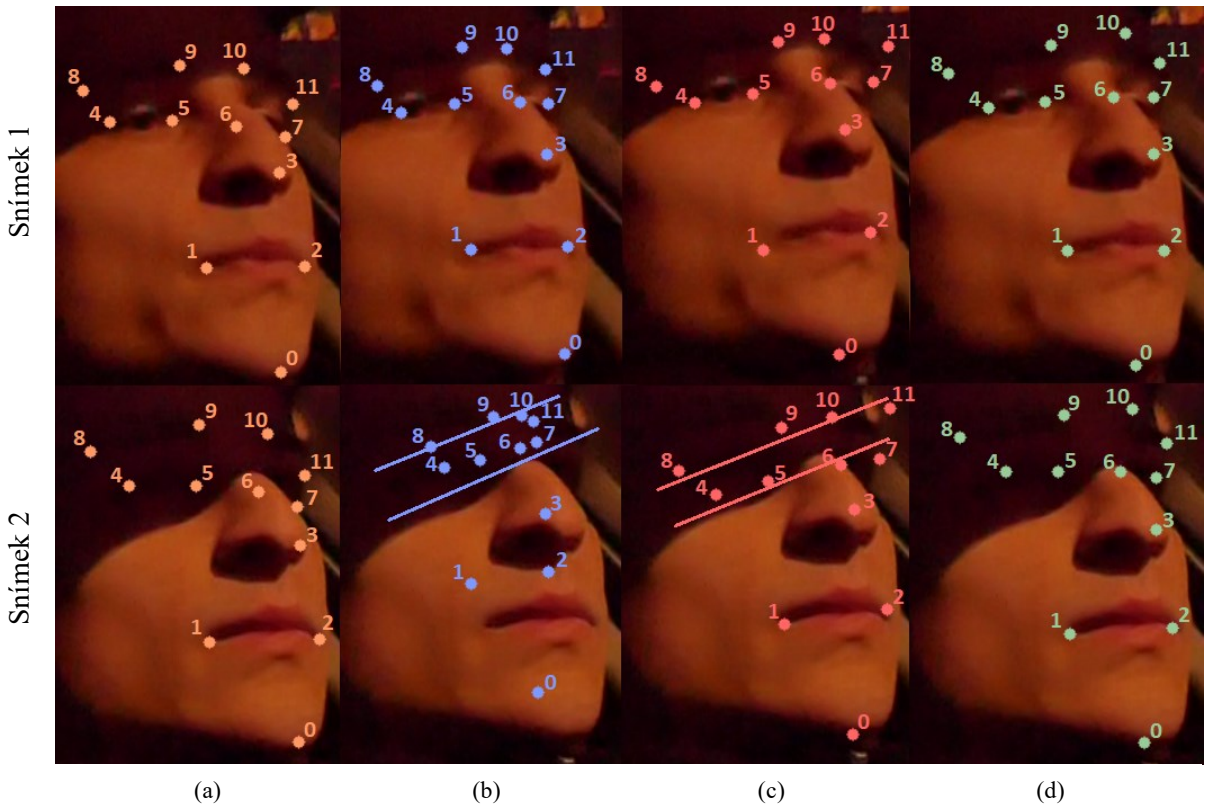
Tabulka 4 – Výsledky testování pro Obrázek 15 (Červeně chyba ≥ 20px).

Model	Chyby predikce obličejových bodů [px]												
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**$\overline{0\text{–}11}$**
WFLW	4,4	3,8	13,8	17,7	**33,8**	16,7	11,3	5,2	**53,5**	10,1	5,6	4,7	15,0
dlib	14,2	3,0	14,0	19,1	**35,0**	15,3	11,0	5,4	**44,0**	**22,4**	9,2	13,9	17,2
2D_FAN	6,4	3,2	8,5	7,3	17,0	5,0	8,2	8,5	19,7	5,0	3,2	12,2	8,7

7.2.5. Ostré stíny

Jedním z problémů doprovázející chybně predikované obličejové body jsou ostré stíny. Ostré stíny vytvářejí ostré předěly, čímž vytváří nové hrany. Ty následně mohou negativně ovlivnit predikce obličejových bodů, respektive modely se jimi mohou nechat zmást. Viz Kapitola 2.1.2.

Problematiku znázorňuje Obrázek 16. Jsou zde v řádcích vyobrazeny celkem dva snímky s predikovanými obličejovými body zvolených metod. Byly zvoleny právě tyto dva snímky, aby bylo možné zobrazit pohyb stínu, stejně tak predikce zvolených metod před a po příchodu stínu. Póza řidiče je téměř totožná na obou prezentovaných snímcích.



Obrázek 16 – Ostrý stín na tváři.
(a) – Manuální anotace; (b) – model WFLW; (c) – model dlib; (d) – model 2D_FAN

Pouhým pozorováním lze usoudit, že model 2D_FAN se, jako jediný z testovaných, nenechal zmást ostrou hranou stínu. Modely WFLW a dlib kopírují sklon stínu (naznačený přímkami). Zejména model WFLW vykazuje velkou chybovost (průměrně 29,6 px pro Snímek 2), pozorovatelnou už z tvaru utvořeného z predikcí. Zatímco model dlib predikuje chybně jen obličejové body v oblasti stínu, model WFLW predikuje chybně obličejové body i mimo stín. Konkrétní vyčíslení chyb uvádí Tabulka 5.

Tabulka 5 – Výsledky testování pro Obrázek 16 (Červeně chyba ≥ 20px).
Pro řádky Δ je chyba definovaná jako $|S2_n - S1_n|$, kde $S2_n$ je chyba n-tého bodu Snímku 2, $S1_n$ je chyba n-tého bodu Snímku 1.

Snímek	Chyby predikce obličejových bodů [px]												
	0	1	2	3	4	5	6	7	8	9	10	11	$\overline{0–11}$
Model WFLW													
1	11,3	4,2	5,7	3,2	16,0	9,5	13,0	13,8	18,8	9,9	5,6	18,0	10,7
2	**36,0**	**31,7**	**48,7**	**24,4**	**25,6**	9,3	**23,0**	**42,8**	**42,8**	11,9	16,1	**43,1**	**29,6**
Δ	24,7	27,5	43,0	21,2	9,6	0,2	10,0	29,0	24,0	2,0	10,5	25,1	18,9
Model dlib													
1	16,6	17,5	10,8	14,1	6,7	1,0	13,5	18,7	16,8	12,4	1,0	**22,4**	12,6
2	13,0	3,0	8,2	16,3	**21,1**	9,1	10,0	**21,5**	**27,7**	16,1	4,1	**34,5**	15,4
Δ	3,6	14,5	2,6	2,2	14,4	8,1	3,5	2,8	10,9	3,7	3,1	12,1	6,8
Model 2D_FAN													
1	11.7	6,1	7,0	3,6	6,1	2,8	10,3	16,0	3,2	3,2	15,3	17,1	8,5
2	12,1	1,4	6,7	7,2	11,2	5,0	9,0	15,1	7,1	2,8	12,2	17,0	8,9
Δ	0,4	4,7	0,3	3,6	5,1	2,2	1,3	0,9	3,9	0,4	3,1	0,1	2,2

7.2.6. Rozmazání

Důsledkem neostrých snímků je na jedné straně menší množství šumu, na druhé straně takový snímek postrádá zejména drobné detaily snímku. V případě velkého rozmazaní může dojít k zániku některých kontur obličeje, což může mít za následek chybné predikce obličejových bodů.

Situaci znázorňuje Obrázek 17, jenž obsahuje dva unikátní snímky. Oba snímky zobrazují velmi podobnou pózu. Liší se ve světelných podmínkách a ostrostí obrazu. Z rozdílu uvedených snímků lze pozorovat vliv ostrosti obrazu na predikce obličejových bodů testovaných metod.

Obrázek 17 – Rozdíl mezi rozmazaným a ostrým snímkem s podobnou pózou.
(a) – Manuální anotace; (b) – model WFLW; (c) – model dlib; (d) – model 2D_FAN

Z pozorování nejlépe vychází model 2D_FAN, jehož většina predikcí vykazují nejmenší odchylku od ručních anotací. Závažnější chybu představují predikce bodů 6 a 11.

Poměrně velkou chybovost predikcí lze pozorovat u zbylých modelů. Z pozorování vyplývá, že predikce (pro Snímek 2) těchto modelů jsou téměř shodné. Sémanticky špatně jsou predikce všech bodů modelu WFLW (Snímek 2 Obrázek 17b). V případě modelu dlib, jsou nejblíže manuálním anotacím, pouze body 4 a 5. Viz Tabulka 6.

Tabulka 6 – Výsledky testování pro Obrázek 17 (Červeně chyba ≥ 20px).
Pro řádky Δ je chyba definovaná jako $|S2_n - S1_n|$, kde $S2_n$ je chyba n-tého bodu Snímku 2, $S1_n$ je chyba n-tého bodu Snímku 1.

Snímek	Chyby predikce obličejových bodů [px]												
	0	1	2	3	4	5	6	7	8	9	10	11	$\overline{0-11}$
Model WFLW													
1	10,1	4,6	11,7	4,8	7,0	3,5	10,4	8,1	16,6	11,9	2,7	19,8	9,3
2	67,6	22,7	63,0	72,1	9,1	18,3	33,2	31,0	27,5	46,6	53,1	28,9	39,4
Δ	57,5	18,1	51,3	67,3	2,1	14,8	22,8	22,9	10,9	34,7	50,4	9,1	30,2
Model dlib													
1	17,8	14,9	18,4	13,0	5,4	1,0	10,6	22,0	12,2	7,8	10,0	40,3	14,4
2	107,3	53,8	75,2	67,2	10,3	10,3	28,0	30,8	32,8	33,7	43,3	25,3	43,2
Δ	89,5	38,9	56,8	54,2	4,9	9,3	17,4	8,8	20,6	25,9	33,3	15,0	31,2
Model 2D_FAN													
1	8,1	3,6	7,1	8,1	4,5	8,5	4,1	10,8	15,5	13,0	4,1	11,3	8,2
2	19,9	7,1	10,0	18,9	8,0	1,4	28,6	18,0	15,0	13,0	12,1	31,0	15,3
Δ	11,8	3,5	2,9	10,8	3,5	7,1	24,5	7,2	0,5	0,0	8,0	19,7	8,3

7.3. Shrnutí

Na závěr testování je nutné uvést celkové výsledky testování podložené číselnými údaji.

Uvedené chyby v tabulkách (Tabulka 7, Tabulka 8 a Tabulka 9) jsou aritmetickým průměrem získaných vzdáleností pro 12 testovaných typů obličejových bodů, viz Kapitola 7.1.3. Uvedené časy predikcí jsou aritmetickým průměrem časů predikcí testovaných vzorků na CPU.

Tabulka 7 – Výsledky testování laboratorních vzorků.

Prostředí	Počet vzorků	Modely predikce					
		WFLW		dlib		2D_FAN	
		Chyba [px]	Čas [s]	Chyba [px]	Čas [s]	Chyba [px]	Čas [s]
Laboratoř	17	7,7	3,979	15,0	0,041	11,0	0,907

Pořízená testovací množina vzorků z reálných podmínek automobilu byla rozšířena o 17 laboratorních snímků, jejichž výsledky testování obsahuje Tabulka 7. Jde o vzorky představující extrémnější situace a pózy, které by bylo nebezpečné zkoušet v reálných podmínkách automobilu např. upadání do mikrospánku apod. Dalším důvodem testování těchto snímků je obohacení testovací množiny o další osoby řidičů, mimiku a gesta. Obličeje jsou snímány téměř výhradně čelně za stabilních světelných podmínek. Ukázkou těchto vzorků je Obrázek 4.

Nejpřesnějších predikcí dosahuje model WFLW s průměrnou chybou $7{,}7\ px$. Velmi přesných predikcí dosahují i zbylé modely 2D_FAN a dlib s chybou $11{,}0\ px$, respektive $15{,}0\ px$.

Z hlediska rychlosti predikce dopadl nejhůře model WFLW s časem $3{,}979\ s$. Ačkoliv jsou predikce tohoto testovaného modelu nejpřesnější, kvůli jeho rychlosti není možné na CPU predikovat obličejové body v reálném čase. Velmi dobré rychlosti $0{,}041\ s$ dosáhl model dlib.

Tabulka 8 – Výsledky testování vzorků z reálných podmínek automobilu.

Světelné podmínky	Počet vzorků	Modely predikce					
		WFLW		dlib		2D_FAN	
		Chyba [px]	Čas [s]	Chyba [px]	Čas [s]	Chyba [px]	Čas [s]
Den – Zataženo	39	20,0	4,184	50,0	0,041	14,3	0,858
Noc	34	17,4	4,068	31,7	0,042	9,6	0,846
Den – Jasno	47	12,1	4,160	32,5	0,041	10,6	0,837
Celkem	120	**16,2**	**4,142**	**37,9**	**0,041**	**11,5**	**0,846**

Na vzorcích z reálných podmínek automobilu je nejrychlejší model dlib, naneštěstí z hlediska přesnosti predikcí je nejhorší. Průměrná chyba je $37{,}9\ px$, což je v porovnání s ostatními testovanými modely závažná hodnota. Z poznatků Kapitoly 7.2 je možné předpokládat, že tato hodnota může být, mimo jiné příčiny, způsobena omezeným rozsahem úhlů póz modelu dlib, viz Kapitola 7.2.1.

Průměrně nejpřesnějších predikcí dosahuje, z testovaných modelů, model 2D_FAN. Jeho průměrnou chybu $11{,}5\ px$ lze interpretovat tím způsobem, že ve většině testovaných vzorků model dosahoval poměrně velké přesnosti. Pro složitější snímky byly ovšem některé predikce více chybové. Průměrný čas predikce $0{,}846\ s$ je z hlediska přesnosti predikcí výborný. Ovšem pro zpracování snímků v reálném čase je nedostatečný.

Za méně přesné lze považovat predikce s chybou $16{,}2\ px$ modelem WFLW. Oproti modelu 2D_FAN se ovšem nejedná o nijak velké zhoršení. Největší nevýhodou modelu je průměrný čas predikcí $4{,}142\ s$. Velmi výkonná grafická karta by ovšem mohla predikce výrazně urychlit.

V porovnání s výsledky z laboratorních podmínek, došlo zejména ke zhoršení přesnosti predikcí modelů WFLW a dlib. Rychlost predikcí je srovnatelná s rychlostí na laboratorních vzorcích.

Tabulka 9 – Celkové výsledky kompletní testované datové množiny.

Počet vzorků	Modely predikce					
	WFLW		dlib		2D_FAN	
	Chyba [px]	Čas [s]	Chyba [px]	Čas [s]	Chyba [px]	Čas [s]
137	15,1	4,122	35,1	0,041	11,5	0,854

Celkově nejpřesnější predikce poskytuje model 2D_FAN. Průměrná rychlost predikcí $0{,}854\ s$ je pro CPU velmi dobrá. Autoři publikace [8] uvádí, že i v případě největší sítě je model 2D_FAN na TITAN X GPU schopen predikovat body za $\sim 0{,}033\ s$, což odpovídá $30\ FPS$. Videa mají obvykle snímkovou frekvenci $\sim 24\ FPS$, z čehož lze usoudit, že model na GPU zvládne predikovat obličejové body v reálném čase.

Model dlib není dle uvedených výsledků predikce vhodný použít v nekontrolovatelných podmínkách jako je např. automobil. Rychlost predikce $0{,}041\ s$ je pro CPU výborná, a téměř odpovídá záměrům autorů, uvést algoritmus predikující obličejové body v řádu jednotek milisekund.

Model WFLW je dle výsledků příliš pomalý, a tak jej není možné použít na CPU pro predikování v reálném čase. Čas $4{,}122\ s$ je poměrně překvapivý, protože publikace [10] uvádí hodnotu $0{,}060\ s$ pro TITAN X GPU. Jde o značný rozdíl v rychlosti i při zanedbání faktu, že hodnota $4{,}122\ s$ platí pro CPU a čas $0{,}060\ s$ pro GPU. Nutno dodat, že výhodou GPU, respektive technologie CUDA je masivní paralelizace výpočtů. Přesnost modelu je jen o $3{,}6\ px$ horší, než přesnost modelu 2D_FAN.

8. Závěr

V rámci teorie bakalářské práce byly představeny tři zvolené metody detekce obličejových bodů. Ke zvoleným metodám byly vytvořeny skripty predikující obličejové body, a to na základě již natrénovaných modelů. Teorie byla také, za účelem zařazení řešené problematiky, doplněna o stručné představení oblastí počítačového vidění a strojového učení.

Dále byla pořízena datová množina z prostředí automobilu. Podařilo se v ní zachytit změny světelných podmínek (zataženo, jasno a noc) i variabilitu póz zachyceného řidiče. Protože variabilita mimiky jednoho řidiče je omezená, byl tento nedostatek v testování minimalizován využitím několika snímků ze simulovaných (laboratorních) podmínek.

V rámci vytvoření testovacího prostředí bylo přistoupeno k vývoji desktopové aplikace LANDMARK. Tím bylo umožněno především vizuální testování predikovaných obličejových bodů, a to včetně těch ručně anotovaných. Další funkcí aplikace bylo vzorkování, respektive možnost sestavení sekvenčních souborů obsahující vzorky pro testování. Aplikace rovněž definovala výstupní formát obličejových bodů příslušných skriptů.

Dále byl vyvinut skript umožňující ke zvoleným vzorkům definovat manuální anotace, vytvořit mapování mezi schématem ručních anotací a schématy příslušných metod, a samozřejmě určit přesnost predikovaných obličejových bodů vůči manuálním anotacím. Rovněž byl vytvořen skript, umožňující získat ze souborů statistik příslušné časy predikcí a vypočítat průměrný čas predikce.

V rámci testování byly rozebrány speciálně snímky představující výzvu pro testované modely. Na závěr byly vyhodnoceny testované modely, jež byly testovány na vzorcích pořízené datové množiny.

Z testování vzešlo, že nejpomalejší je na CPU model WFLW s časem $4{,}122\ s$, ovšem mnohem lepších časů by mělo být možné dosáhnout na GPU. Poměrně rychlý je na CPU model 2D_FAN s časem $0{,}854\ s$. Nejrychlejší je model dlib s výborným časem $0{,}041\ s$.

Z hlediska přesnosti predikcí dosahuje nejmenší chyby model 2D_FAN s průměrnou chybou $11{,}5\ px$. O $3{,}6\ px$ horší přesnosti dosahuje model WFLW s průměrnou chybou $15{,}1\ px$. Průměrně největší chyby predikcí $35{,}1\ px$ dosahuje model dlib.

Literatura

[1] Pajankar, Ashwin. *Raspberry Pi Computer Vision Programming.* 1. Birmingham : Packt Publishing Ltd., 2015. 978-1-78439-828-6.

[2] DAVIES, E. R. *Computer and Machine Vision: Theory, Algorithms, Practicalities.* 4. Waltham : ELSEVIER, 2012. ISBN: 978-0-12-386908-1.

[3] OpenCV. *OpenCV.* [Online] OpenCV. [Citace: 20. únor 2020.] https://opencv.org/about/.

[4] King, Davis E. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research.* 2009. 10, stránky 1755-1758.

[5] Patterson, Josh a Gibson, Adam. *Deep Learning - A Practitioner's Approach.* 1. Sebastopo : O'Reilly Media, Inc., 2017. ISBN 9781491914250.

[6] Bell, Jason. *Machine Learning: Hands-On for Developers and Technical Professionals.* Indianapolis : John Wiley & Sons, Inc., 2015. stránky 45-49. ISBN: 978-1-118-88906-0.

[7] *300 Faces in-the-Wild Challenge: The First Facial Landmark Localization Challenge.* Sagonas, Christos, a další. Prosinec 2013, stránky 397-403.

[8] Bulat, Adrian a Tzimiropoulos, Georgios. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). *International Conference on Computer Vision.* 2017.

[9] *Interactive Facial Feature Localization.* Le, Vuong, a další. Říjen 2012.

[10] Wayne, Wu, a další. Look at Boundary: A Boundary-Aware Face Alignment Algorithm. *CVPR.* Červen 2018.

[11] GitHub. *Which datasets have you used?* [Online] 5. Září 2018. [Citace: 23. Březen 2020.] https://github.com/wywu/LAB/issues/13.

[12] Preprocessing alignment using 71pts. *GitHub.* [Online] 23. Srpen 2018. [Citace: 12. Březen 2020.] https://github.com/wywu/LAB/issues/9.

[13] How to evaluate on new face data. *GitHub.* [Online] 6. Únor 2019. [Citace: 29. Únor 2020.] https://github.com/wywu/LAB/issues/32.

[14] Easy-LAB. *GitHub.* [Online] [Citace: 12. Březen 2020.] https://github.com/HandsomeHans/Easy-LAB.

[15] Friedman, Jerome. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics.* Říjen 2001, 29, stránky 1189-1232.

[16] Kazemi, Vahid a Sullivan, Josephine. One Millisecond Face Alignment with an Ensemble of Regression Trees. *CVPR.* 2014.

[17] Real-Time Face Pose Estimation . *dlib C++ Library.* [Online] [Citace: 8. Března 2020.] http://blog.dlib.net/2014/08/real-time-face-pose-estimation.html.

[18] ibug_300W_large_face_landmark_dataset. *Dlib.* [Online] [Citace: 8. Březen 2020.] http://dlib.net/files/data/ibug_300W_large_face_landmark_dataset.tar.gz.

[19] train_shape_predictor_ex.cpp. *Dlib.* [Online] [Citace: 8. Březen 2020.] http://dlib.net/train_shape_predictor_ex.cpp.html.

[20] face_landmark_detection_ex. *Dlib.* [Online] [Citace: 8. Březen 2020.] http://dlib.net/face_landmark_detection_ex.cpp.html.

[21] Decision Tree Classification Algorithm. *Java T Point.* [Online] [Citace: 12. Března 2020.] https://static.javatpoint.com/tutorial/machine-learning/images/decision-tree-classification-algorithm.png.

[22] Helen dataset. *Helen dataset.* [Online] [Citace: 31. Březen 2020.] http://www.ifp.illinois.edu/~vuongle2/helen/.

[23] Face Alignment Across Large Poses: A 3D Solution. *[Data].* [Online] [Citace: 31. Březen 2020.] http://www.cbsr.ia.ac.cn/users/xiangyuzhu/projects/3DDFA/main.htm.

[24] Look at Boundary: A Boundary-Aware Face Alignment Algorithm. *WFLW.* [Online] [Citace: 31. Březen 2020.] https://wywu.github.io/projects/LAB/WFLW.html.

[25] A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. *Towards Data Science.* [Online] Medium, 15. Prosinec 2018. [Citace: 4. Duben 2020.] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

I. Příloha

- testované vzorky z reálných podmínek automobilu
- zdrojové soubory:
    - aplikace LANDMARK,
    - evaluačních skriptů testovaných metod,
    - skriptu pro zpracování statistik,
    - skriptu pro vytváření manuálních anotací